

Statistical Methods - Homework #2

1. Consider the logistic regression model

$$\log \frac{p_i}{1 - p_i} = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad y_i \sim \text{Bernoulli}(p_i) \text{ independently, } i = 1, \dots, n.$$

(A) Show that

$$p_i = \frac{1}{1 + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta})}$$

Also, find $\frac{\partial p_i}{\partial \boldsymbol{\beta}}$.

Solution: Exponentiating both sides of the logit equation gives $\frac{p_i}{1 - p_i} = e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$, so $p_i = (1 - p_i)e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$. Solving for

p_i :

$$p_i = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}} = \boxed{\frac{1}{1 + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta})}}.$$

For the derivative, note that $1 - p_i = \frac{1}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}$ and $\frac{\partial(\mathbf{x}_i^\top \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{x}_i$. By the chain rule,

$$\frac{\partial p_i}{\partial \boldsymbol{\beta}} = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}})^2} \mathbf{x}_i = \boxed{p_i(1 - p_i) \mathbf{x}_i}.$$

(B) The likelihood $L(\boldsymbol{\beta})$ and the log-likelihood $\ell(\boldsymbol{\beta})$ of $\boldsymbol{\beta}$ are

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1 - y_i}, \quad \ell(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \log p_i + (1 - y_i) \log(1 - p_i)$$

respectively. Find $S(\boldsymbol{\beta}) = \frac{\partial \ell}{\partial \boldsymbol{\beta}}$ and $\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}$.

Solution: Differentiating $\ell(\boldsymbol{\beta})$ using the chain rule and the result of (A),

$$\begin{aligned} S(\boldsymbol{\beta}) &= \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i} \right) \frac{\partial p_i}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^n \frac{y_i(1 - p_i) - (1 - y_i)p_i}{p_i(1 - p_i)} \cdot p_i(1 - p_i) \mathbf{x}_i \\ &= \boxed{\sum_{i=1}^n (y_i - p_i) \mathbf{x}_i}. \end{aligned}$$

Differentiating once more with respect to $\boldsymbol{\beta}^\top$ and using $\frac{\partial p_i}{\partial \boldsymbol{\beta}^\top} = p_i(1 - p_i) \mathbf{x}_i^\top$,

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = - \sum_{i=1}^n p_i(1 - p_i) \frac{\partial p_i}{\partial \boldsymbol{\beta}^\top} \cdot \frac{1}{p_i(1 - p_i)} \cdot p_i(1 - p_i) = \boxed{- \sum_{i=1}^n p_i(1 - p_i) \mathbf{x}_i \mathbf{x}_i^\top}.$$

(C) The maximum likelihood estimator (MLE) $\hat{\beta}$ solves $S(\hat{\beta}) = 0$. Find the asymptotic distribution of the MLE $\hat{\beta}$.

Solution: The Fisher information matrix is

$$\mathcal{I}(\beta) = -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \beta \partial \beta^\top} \right] = \sum_{i=1}^n p_i (1 - p_i) \mathbf{x}_i \mathbf{x}_i^\top.$$

By standard maximum likelihood theory, under regularity conditions,

$$\boxed{\hat{\beta} \xrightarrow{a} N(\beta, \mathcal{I}(\beta)^{-1})},$$

i.e., $\hat{\beta}$ is asymptotically unbiased with variance-covariance matrix $\mathcal{I}(\beta)^{-1} = \left(\sum_{i=1}^n p_i (1 - p_i) \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1}$. In

practice, $\mathcal{I}(\beta)$ is replaced by the observed information $\mathcal{I}(\hat{\beta})$.

(D) Consider the hypothesis test:

$$H_0 : \boldsymbol{\beta} = \mathbf{0}, \quad H_A : \boldsymbol{\beta} \neq \mathbf{0}.$$

Show that the Wald test statistic is

$$W = \sum_{i=1}^n \hat{p}_i(1 - \hat{p}_i) \left(\log \frac{\hat{p}_i}{1 - \hat{p}_i} \right)^2,$$

and the likelihood ratio statistic is

$$\Lambda = 2 \sum_{i=1}^n \{y_i \log(2\hat{p}_i) + (1 - y_i) \log(2(1 - \hat{p}_i))\},$$

where $\hat{p}_i = p_i(\hat{\boldsymbol{\beta}})$. What are the null distribution of these test statistics?

Solution: *Wald statistic.* The general Wald statistic for $H_0 : \boldsymbol{\beta} = \mathbf{0}$ is $W = \hat{\boldsymbol{\beta}}^\top \mathcal{I}(\hat{\boldsymbol{\beta}}) \hat{\boldsymbol{\beta}}$. Using

$$\mathcal{I}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^n \hat{p}_i(1 - \hat{p}_i) \mathbf{x}_i \mathbf{x}_i^\top,$$

$$W = \hat{\boldsymbol{\beta}}^\top \left(\sum_{i=1}^n \hat{p}_i(1 - \hat{p}_i) \mathbf{x}_i \mathbf{x}_i^\top \right) \hat{\boldsymbol{\beta}} = \sum_{i=1}^n \hat{p}_i(1 - \hat{p}_i) (\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})^2.$$

Since $\mathbf{x}_i^\top \hat{\boldsymbol{\beta}} = \log \frac{\hat{p}_i}{1 - \hat{p}_i}$ by definition of \hat{p}_i ,

$$W = \sum_{i=1}^n \hat{p}_i(1 - \hat{p}_i) \left(\log \frac{\hat{p}_i}{1 - \hat{p}_i} \right)^2.$$

Likelihood ratio statistic. Under $H_0 : \boldsymbol{\beta} = \mathbf{0}$, all $p_i = \frac{1}{2}$, so

$$\ell(\mathbf{0}) = \sum_{i=1}^n [y_i \log \frac{1}{2} + (1 - y_i) \log \frac{1}{2}] = -n \log 2.$$

Therefore,

$$\begin{aligned} \Lambda &= 2 [\ell(\hat{\boldsymbol{\beta}}) - \ell(\mathbf{0})] \\ &= 2 \sum_{i=1}^n [y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i)] + 2n \log 2 \\ &= 2 \sum_{i=1}^n [y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i) + y_i \log 2 + (1 - y_i) \log 2] \\ &= 2 \sum_{i=1}^n \{y_i \log(2\hat{p}_i) + (1 - y_i) \log(2(1 - \hat{p}_i))\}. \end{aligned}$$

Null distributions. Let p be the dimension of $\boldsymbol{\beta}$. Under H_0 , both W and Λ follow asymptotically a chi-squared distribution with p degrees of freedom:

$$W \xrightarrow{d} \chi^2(p), \quad \Lambda \xrightarrow{d} \chi^2(p).$$

2. Consider the logistic regression model with a single binary covariate $x_i \in \{0, 1\}$:

$$\log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_i, \quad y_i \sim \text{Bernoulli}(p_i) \text{ independently, } i = 1, \dots, n.$$

The data are summarized in the following contingency table:

		Y		Total
		0	1	
X	0	n_{00}	n_{01}	$n_{00} + n_{01} = n_0$
	1	n_{10}	n_{11}	$n_{10} + n_{11} = n_1$
Total		$n_{00} + n_{10}$	$n_{01} + n_{11}$	$n_{00} + n_{01} + n_{10} + n_{11}$

(A) For subjects with $x_i = 0$, $p_i = p(0)$; for $x_i = 1$, $p_i = p(1)$, where $p(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$ and $p(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$.

Express the odds ratio(OR)

$$\text{OR} = \frac{p(1)/(1 - p(1))}{p(0)/(1 - p(0))}$$

in terms of β_0 and β_1 .

Solution: From the logistic model, the odds for each group are

$$\frac{p(0)}{1 - p(0)} = e^{\beta_0}, \quad \frac{p(1)}{1 - p(1)} = e^{\beta_0 + \beta_1}.$$

Therefore,

$$\text{OR} = \frac{p(1)/(1 - p(1))}{p(0)/(1 - p(0))} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = \boxed{e^{\beta_1}}.$$

Hence $\beta_1 = \log(\text{OR})$ is the log odds ratio of $Y = 1$ for $X = 1$ versus $X = 0$.

(B) Denote $n_0 = n_{00} + n_{01}$ and $n_1 = n_{10} + n_{11}$. Using the score equations $S(\beta) = \mathbf{0}$ from 1.(B), show that the MLEs of $p(0)$ and $p(1)$ are

$$\hat{p}(0) = \frac{n_{01}}{n_0}, \quad \hat{p}(1) = \frac{n_{11}}{n_1}.$$

Solution: Since all subjects with $x_i = 0$ share the same probability $p(0)$, and similarly for $p(1)$, the score equations $S(\beta) = \mathbf{0}$ from 1.(B) decouple by group:

$$\begin{aligned} \sum_{i: x_i=0} (y_i - p(0)) = 0 &\Rightarrow n_{01} - n_0 p(0) = 0 \Rightarrow \boxed{\hat{p}(0) = \frac{n_{01}}{n_0}}, \\ \sum_{i: x_i=1} (y_i - p(1)) = 0 &\Rightarrow n_{11} - n_1 p(1) = 0 \Rightarrow \boxed{\hat{p}(1) = \frac{n_{11}}{n_1}}. \end{aligned}$$

(C) Using the result of (B), find the closed form MLEs $\hat{\beta}_0$ and $\hat{\beta}_1$.

Solution: Inverting the logit $\beta_0 = \log \frac{p(0)}{1-p(0)}$ at the MLE $\hat{p}(0)$,

$$\hat{\beta}_0 = \log \frac{\hat{p}(0)}{1-\hat{p}(0)} = \boxed{\log \frac{n_{01}}{n_{00}}}.$$

Since $\beta_0 + \beta_1 = \log \frac{p(1)}{1-p(1)}$, we have $\hat{\beta}_0 + \hat{\beta}_1 = \log \frac{n_{11}}{n_{10}}$, and therefore

$$\hat{\beta}_1 = \log \frac{n_{11}}{n_{10}} - \log \frac{n_{01}}{n_{00}} = \boxed{\log \frac{n_{11} n_{00}}{n_{10} n_{01}}}.$$

Note that $\hat{\beta}_1 = \log \widehat{OR}$, consistent with part (A).

(D) Let $q(j) = 1 - p(j)$ for $j = 0, 1$. Using the result of 1.(B) with $\mathbf{x}_i = (1, x_i)^\top$, show that the Fisher information matrix is

$$\mathcal{I}(\boldsymbol{\beta}) = \begin{pmatrix} n_0 p(0)q(0) + n_1 p(1)q(1) & n_1 p(1)q(1) \\ n_1 p(1)q(1) & n_1 p(1)q(1) \end{pmatrix}.$$

Then find the asymptotic variance of $\hat{\beta}_1$, and evaluate it at the MLE.

Solution: From 1.(B), $\mathcal{I}(\boldsymbol{\beta}) = -\mathbf{E}[\partial^2 \ell / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top] = \sum_{i=1}^n p_i(1-p_i)\mathbf{x}_i\mathbf{x}_i^\top$. Since $p_i = p(0)$ for $x_i = 0$ and $p_i = p(1)$ for $x_i = 1$,

$$\mathcal{I}(\boldsymbol{\beta}) = \sum_{i: x_i=0} p(0)q(0) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \sum_{i: x_i=1} p(1)q(1) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} n_0 p(0)q(0) + n_1 p(1)q(1) & n_1 p(1)q(1) \\ n_1 p(1)q(1) & n_1 p(1)q(1) \end{pmatrix}.$$

By the formula for the inverse of a 2×2 matrix, the $(2, 2)$ entry of $\mathcal{I}(\boldsymbol{\beta})^{-1}$ gives

$$\text{Var}(\hat{\beta}_1) = \frac{n_0 p(0)q(0) + n_1 p(1)q(1)}{\det \mathcal{I}(\boldsymbol{\beta})} = \frac{n_0 p(0)q(0) + n_1 p(1)q(1)}{n_0 p(0)q(0) \cdot n_1 p(1)q(1)} = \boxed{\frac{1}{n_0 p(0)q(0)} + \frac{1}{n_1 p(1)q(1)}}.$$

Evaluated at the MLE, noting $n_j \hat{p}(j) \hat{q}(j) = \frac{n_{j0} n_{j1}}{n_{j0} + n_{j1}}$ for $j = 0, 1$,

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{1}{n_0 \hat{p}(0) \hat{q}(0)} + \frac{1}{n_1 \hat{p}(1) \hat{q}(1)} = \boxed{\frac{1}{n_{00}} + \frac{1}{n_{01}} + \frac{1}{n_{10}} + \frac{1}{n_{11}}}.$$

(E) Find a test statistic and its null distribution for the following hypothesis test:

$$H_0 : \beta_1 = 0, \quad H_A : \beta_1 \neq 0.$$

Solution: By the asymptotic normality of the MLE from 1.(C), under $H_0 : \beta_1 = 0$,

$$Z = \frac{\hat{\beta}_1}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}} = \frac{\log \frac{n_{11} n_{00}}{n_{10} n_{01}}}{\sqrt{\frac{1}{n_{00}} + \frac{1}{n_{01}} + \frac{1}{n_{10}} + \frac{1}{n_{11}}}} \xrightarrow{d} N(0, 1).$$

Equivalently, the Wald statistic $W = Z^2$ satisfies

$$\boxed{W = \frac{\hat{\beta}_1^2}{\widehat{\text{Var}}(\hat{\beta}_1)} \xrightarrow{d} \chi^2(1)} \quad \text{under } H_0.$$

3. (Linear Discriminant Analysis) Show that finding $k \in \{1, \dots, K\}$ that maximizes

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu_k)^2\right\}}{\sum_{\ell=1}^K \pi_\ell \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu_\ell)^2\right\}}$$

is equivalent to finding $k \in \{1, \dots, K\}$ that maximizes

$$\delta_k(x) = x \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k).$$

Also, when $K = 2$, show that this is equivalent to choosing $k = 1$ if

$$\frac{(\mu_2 - \mu_1)}{\sigma^2} \left(x - \frac{\mu_1 + \mu_2}{2} \right) < \log\left(\frac{\pi_1}{\pi_2}\right).$$

Solution:

Since the denominator of $p_k(x)$ does not depend on k , maximizing $p_k(x)$ is equivalent to maximizing the numerator.

Taking logs,

$$\log(\pi_k) - \frac{1}{2} \log(2\pi) - \log(\sigma) - \frac{(x - \mu_k)^2}{2\sigma^2}.$$

Dropping constants in k and expanding gives

$$-\frac{x^2}{2\sigma^2} + \frac{\mu_k}{\sigma^2} x - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k).$$

The term $-x^2/(2\sigma^2)$ does not depend on k , so the maximizer is

$$\delta_k(x) = \frac{\mu_k}{\sigma^2} x - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k).$$

For $K = 2$, choose class 1 if $\delta_1(x) > \delta_2(x)$, i.e.

$$\left(\frac{\mu_1 - \mu_2}{\sigma^2}\right) x - \frac{\mu_1^2 - \mu_2^2}{2\sigma^2} + \log\left(\frac{\pi_1}{\pi_2}\right) > 0.$$

Using $\mu_1^2 - \mu_2^2 = (\mu_1 - \mu_2)(\mu_1 + \mu_2)$ and rearranging yields

$$\frac{\mu_2 - \mu_1}{\sigma^2} \left(x - \frac{\mu_1 + \mu_2}{2} \right) < \log\left(\frac{\pi_1}{\pi_2}\right).$$

4. Consider the following confusion matrix and answer the following questions.

		True	
		Negative	Positive
Predicted	Negative	75	20
	Positive	25	80

(1) Compute the False Positive Rate (FPR).

Solution:

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN}) = 25 / (25 + 75) = 25/100 = 1/4 \text{ (25\%)}$$

(2) Compute the False Negative Rate (FNR).

Solution:

$$\text{FNR} = \text{FN} / (\text{FN} + \text{TP}) = 20 / (20 + 80) = 20/100 = 1/5 \text{ (20\%)}$$

(3) Compute the Sensitivity (True Positive Rate).

Solution:

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) = 80 / (80 + 20) = 80/100 = 4/5 \text{ (80\%)}$$

(4) Compute the Specificity (True Negative Rate).

Solution:

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) = 75 / (75 + 25) = 75/100 = 3/4 \text{ (75\%)}$$