Longitudinal Data Analysis Group Project

Simulations of IRLS and Fisher Scoring for logistic regression

Chloe Mattila, Courtney Jacks, Yao Xin

November 2024

1 Introduction

This report develops the algorithm of Iterative Re-weighted Least Squares(IRLS) and Fisher Scoring for estimating the beta(?) in logistic regression with a logit link, and Fisher Scoring for logistic regression with a probit link.

The math and algorithm are based on Dr. Brian Neelon's lecture of *Likelihood-based Inference for Univariate GLMs.*

2 Logistic regression with logit/probit link

2.1 Exponential family likelihood of logistic regression

Given that each *i.i.d.* binary $Y_i \sim \text{Bernoulli}(\pi_i)$, we have:

$$f(y_i;\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}, \quad y_i \in \{0,1\}, 0 \leq \pi \leq 1$$

which can be rewritten as:

$$f(y_i;\pi_i) = \exp\left\{y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log(1-\pi_i)\right\}$$

Let the canonical parameter $\theta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)$, then:

$$f(y_i;\pi_i) = \exp\left\{y_i\theta_i - \log(1+e^{\theta_i})\right\}$$

The likelihood is:

$$L(\pi_i;y_i) = \exp\left\{y_i\theta_i - \log(1+e^{\theta_i})\right\}$$

2.2 Using a logit link

The systematic component is given by $\eta = \mathbf{X}\beta$, when using the canonical link (the logit function), the canonical parameter $\theta_i = \eta_i = g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$, $A(\eta_i) = \log(1+e^{\eta_i})$, and the dispersion parameter $\phi = 1$.

Thus, the mean μ_i is given by:

$$\mu_i = E(Y_i|X) = A'(\eta_i) = \frac{e^{\eta_i}}{1+e^{\eta_i}} \quad (=\pi_i)$$

The variance $\mathrm{Var}(\mu_i)$ is:

$$\mathrm{Var}(\mu_i) = \pi_i(1-\pi_i)$$

2.3 Using a Probit link

When using the Probit function as link function, the systematic component $\eta = \Phi^{-1}(\pi)$, where Φ^{-1} is the inverse of the normal CDF (the probit function).

3 Fisher Scoring

3.1 General algorithm

$$\hat{\beta} = \beta + \mathcal{I}^{-1}(\beta)\mathcal{U}(\beta)$$

where

$$\begin{cases} \mathcal{U}(\beta) = \mathbf{D}^T\mathbf{V}^{-1}(\mathbf{y}-\mu) \\ \mathcal{I}^{-1}(\beta) = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1} \end{cases}$$

$$\mathbf{D}_{n\times p} = \frac{\partial\mu}{\partial\beta} = \frac{\partial\mu}{\partial\eta}\frac{\partial\eta}{\partial\beta} = \frac{\partial\mu}{\partial\eta}\frac{\partial(\mathbf{X}\beta)}{\partial\beta} = \mathbf{D}^*\mathbf{X},$$

where $\mathbf{D}^* = \mathrm{diag}\left[\frac{\partial\mu_i}{\partial\eta_i}\right]$.

Taking $\mathbf{D} = \mathbf{D}^*\mathbf{X}$ into $\mathcal{I}^{-1}(\beta)$,

$$\begin{aligned} \mathcal{I}^{-1}(\beta) &= \left(\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D}\right)^{-1} \\ &= \left(\mathbf{X}^T\mathbf{D}^*\mathbf{V}^{-1}\mathbf{D}^*\mathbf{X}\right)^{-1} \\ &= \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}, \end{aligned}$$

where $\mathbf{W} = \mathbf{D}^*\mathbf{V}^{-1}\mathbf{D}^*$.

Now, using *property 2* (slide 26), which says that if $\mathrm{V}(\boldsymbol{Y}) = \phi\mathbf{v}(\mu)$ (ϕ is the dispersion parameter from the exponential dispersion family) is correctly specified and the EF-type regularity conditions hold, then

$$\mathrm{Cov}(\widehat{\beta}) = \mathcal{I}^{-1}(\boldsymbol{\beta}) = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}$$

3.2 Fisher Scoring on logistic regression with logit link

Based on the data, we can obtain $\eta = \mathbf{X}\beta$, and thus $\mu = \pi = \frac{\exp(\eta)}{1+\exp(\eta)}$.
Under the canonical link,

$$\begin{cases} \mathcal{U}(\beta) = \mathbf{X}^T(\mathbf{y} - \mu)/\phi \\ \mathcal{I}^{-1}(\beta) = \phi(\mathbf{X}^T v(\mu)\mathbf{X})^{-1} \end{cases}$$

where $v(\mu) = diag[\pi_i(1 - \pi_i)]$, and $\phi = 1$.

Fisher scoring is performed by plugging these in

$$\hat{\beta} = \beta + \mathcal{I}^{-1}(\beta)\mathcal{U}(\beta)$$

.

For $\mathrm{Cov}(\widehat{\beta})$, under the canonical link,

$$\mathrm{Cov}(\widehat{\beta}) = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1} = \phi[\mathbf{X}^T v(\mu)\mathbf{X}]^{-1} = \{\mathbf{X}^T diag[\mu_i(1 - \mu_i)]\mathbf{X}\}^{-1}$$

3.3 Fisher Scoring on logistic regression with probit link

Since $\boldsymbol{\mu} = \Phi(\boldsymbol{\eta})$, then it follows that

$$\mathbf{D}^* = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} = \phi(\boldsymbol{\eta}) = \phi(\mathbf{X}\beta),$$

where Φ is the standard normal CDF and ϕ is the standard normal pdf.
Thus,

$$\mathbf{D} = \mathbf{D}^*\mathbf{X} = \phi(\boldsymbol{\eta})\mathbf{X}$$

Fisher Scoring is performed by plugging in $\mathbf{D}$, $\mu = \Phi(\mathbf{X}\beta)$ and $V = \phi diag[v(\mu)] = diag[\mu_i(1 - \mu_i)]$ into

$$\begin{cases} \mathcal{U}(\beta) = \mathbf{D}^T\mathbf{V}^{-1}(\mathbf{y} - \mu) \\ \mathcal{I}^{-1}(\beta) = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1} \end{cases}$$

Fisher scoring is performed by plugging these in updating

$$\hat{\beta} = \beta + \mathcal{I}^{-1}(\beta)\mathcal{U}(\beta)$$

And $\mathrm{Cov}(\widehat{\beta}) = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}$

4 IRLS

4.1 General algorithm

Based on the Fisher scoring set-up, the estimation procedure iterates the following two steps, until $max|\hat{\beta}_{new} - \hat{\beta}_{old}| < \epsilon \approx 0$.

[1] With given value of β, update

$$\begin{cases} \mathbf{Z} = \mathbf{X}\beta + \mathbf{D}^{*-1}(\mathbf{y} - \mu) \\ \mathbf{W} = \mathbf{D}^{*}\mathbf{V}^{-1}\mathbf{D}^{*} \end{cases}$$

[2] With given values of $\mathbf{Z}$ and $\mathbf{W}$, update $\hat{\beta}$ using Weighted Least Squares (WLS),

$$\beta = \beta + \mathcal{I}^{-1}(\beta)\mathcal{U}(\beta) = (\mathbf{X^T W X})^{-1}\mathbf{X^T W Z}$$

And with the final estimated W,

$$Cov(\hat{\beta}) = (X^T W X)^{-1}$$

4.2 IRLS on logistic regression with logit link

Based on the data, we can obtain $\eta = \mathbf{X}\beta$, and thus $\mu = \pi = \frac{\exp(\eta)}{1+\exp(\eta)}$.
For a canonical link,

$$\mathbf{D}^* = diag[v(\mu_i)]$$

$$\mathbf{W} = diag[v(\mu_i)]diag[\phi v(\mu_i)]^{-1}diag[v(\mu_i)] = diag[v(\mu_i)]/\phi = diag[v(\mu_i)]$$

Based on section 2.2, $v(\mu_i) = \pi_i(1 - \pi_i)$, thus

$$\mathbf{D}^* = \mathbf{W} = diag[v(\mu_i)] = diag[\pi_i(1 - \pi_i)]$$

Thus, step [1] of IRLS updates:

$$\begin{cases} \mathbf{Z} = \mathbf{X}\beta + diag[\pi_i(1 - \pi_i)]^{-1}(\mathbf{y} - \frac{\exp(\eta)}{1+\exp(\eta)}) \\ \mathbf{W} = diag[\pi_i(1 - \pi_i)] \end{cases}$$

then step [2] updates β with

$$\beta = (\mathbf{X^T W X})^{-1}\mathbf{X^T W Z}$$

and $Cov(\hat{\beta}) = (X^T W X)^{-1}$, which is the same as in 3.2.

5 Simulation in R

See Rmd report.