

Scaled logit model at true baseline of 1

Arseniy Khvorov

12/11/2019

All relevant code is at github.com/khvorov45/sclr-boundary

1 Parameterisation

The original parameterisation

$$P(Y_i = 1) = \frac{\lambda}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta})} \quad (1)$$

Where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \dots & X_{k,1} \\ 1 & X_{1,2} & X_{2,2} & \dots & X_{k,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1,n} & X_{2,n} & \dots & X_{k,n} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}$$

Constrains λ to $[0, 1]$. This presents two problems:

1. Optimisation algorithms may produce out-of-bounds guesses for λ .
2. Normal approximation for the estimate of λ may be inappropriate and the bounds of the confidence interval may go out of $[0, 1]$.

To address this, I reparameterised the model as follows

$$P(Y_i = 1) = \frac{\exp(\theta)}{(1 + \exp(\theta))(1 + \exp(\mathbf{X}_i \boldsymbol{\beta}))} \quad (2)$$

Where

$$\theta = \log\left(\frac{\lambda}{1 - \lambda}\right)$$

2 Likelihood

To explore the likelihood of the scaled logit model when it is fit to data with a true baseline of 1, I simulated one dataset from the model

$$P(Y_i = 1) = \frac{1}{1 + \exp(-5 + 2X_i)}$$

Where X represents the log HI titre measurement.

The perfect model for this data is the standard logistic model

$$P(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

Except that the β parameters have the opposite sign to their equivalents in the simulation model (i.e. the true value of β_0 is 5 (not -5) and the true value of β_1 is -2).

Fitting the standard logistic model to the data gave the following maximum likelihood estimates

$$\hat{\beta}_0 = 4.46 \quad \hat{\beta}_1 = -1.80$$

Figure 1 shows the log-likelihood of scaled logit model

$$l(\theta, \beta) = \sum_i y_i \theta - \log(1 + \exp(\mathbf{X}_i \beta)) - \log(1 + \exp(\theta)) + (1 - y_i) \log(1 + \exp(\mathbf{X}_i \beta) (1 + \exp(\theta))) \quad (3)$$

as a function of θ conditional on $\beta_0 = -4.46$ and $\beta_1 = 1.80$

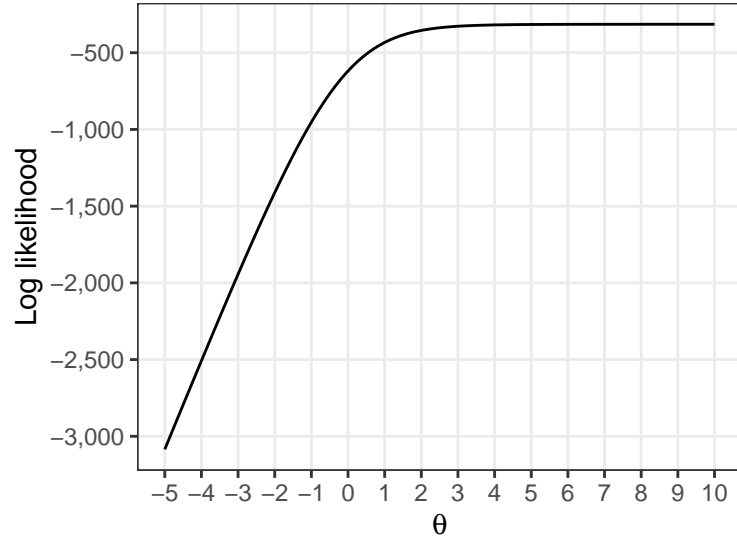


Figure 1: Likelihood of the scaled logit model as a function of θ conditional on the β parameters set to their maximum likelihood estimates obtained from fitting the standard logistic model.

The reason why the likelihood flattens out is that the true baseline probability λ is 1 and any value of baseline log-odds θ above around 5 corresponds to the baseline probability close to 1.

3 Optimisation

Whatever algorithm is applied to the likelihood in Figure 1, it will start walking toward larger and larger values of θ . The problem is that first (and second) derivatives of log-likelihood

$$\begin{aligned} \begin{bmatrix} \frac{dl}{d\lambda} \\ \frac{dl}{d\beta_j} \end{bmatrix} &= \begin{bmatrix} \sum_i y_i - \frac{\exp(\theta)}{1+\exp(\theta)} + \frac{(1-y_i)\exp(\mathbf{X}_i\boldsymbol{\beta})\exp(\theta)}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})(1+\exp(\theta))} \\ \sum_i x_{j,i} \left(-\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})} + \frac{(1-y_i)\exp(\mathbf{X}_i\boldsymbol{\beta})(1+\exp(\theta))}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})(1+\exp(\theta))} \right) \end{bmatrix} \\ \frac{dl}{d\lambda} &\left[\sum_i -\frac{\exp(\theta)}{(1+\exp(\theta))^2} + \frac{\frac{dl}{d\lambda}(1+\exp(\mathbf{X}_i\boldsymbol{\beta}))\exp(\mathbf{X}_i\boldsymbol{\beta})\exp(\theta)}{\left(1+\exp(\mathbf{X}_i\boldsymbol{\beta})(1+\exp(\theta))\right)^2} \right. \\ \frac{dl}{d\beta_j} &\left[\begin{array}{c} \sum_i x_{j,i} \left(\frac{\frac{dl}{d\beta_j}(1-y_i)\exp(\mathbf{X}_i\boldsymbol{\beta})\exp(\theta)}{\left(1+\exp(\mathbf{X}_i\boldsymbol{\beta})(1+\exp(\theta))\right)^2} \right) \\ \sum_i x_{j,i} \left(-\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{(1+\exp(\mathbf{X}_i\boldsymbol{\beta}))^2} + \frac{(1-y_i)(1+\exp(\theta))\exp(\mathbf{X}_i\boldsymbol{\beta})}{\left(1+\exp(\mathbf{X}_i\boldsymbol{\beta})(1+\exp(\theta))\right)^2} \right) \end{array} \right] \end{aligned}$$

have e^θ terms and at large values of θ it becomes impossible to calculate some of them.

To overcome this problem I've implemented the gradient ascent algorithm (in addition to Newton-Raphson) which stops when the latest 20% of iterations incremented the log-likelihood by less than 0.001.

Whenever Newton-Raphson fails to converge, gradient ascent will attempt to find the optimal set of estimates. This set can be used to calculate the maximum log-likelihood value of the scaled logit model and compare it to the maximum likelihood value of the standard logistic model fit to the same data. If the two values are (almost) the same then the addition of the baseline log-odds parameter θ does not result in a better fit, therefore the baseline probability should be fixed to 1 by fitting (and inferring from) the standard logistic model.

I've implemented a function `check_baseline` into the `sclr` package which fits the scaled logit and the logistic models and performs the likelihood ratio test to aid model selection.