STAT5514: Topics in Regression

Inyoung Kim
Outline

1 Fitting GLMs
Fitting GLMS

- We study how to find the maximum likelihood estimator $\hat{\beta}$ of GLM parameters.
- The likelihood equations are usually nonlinear in $\hat{\beta}$.
- We will describe a general-purpose iterative method for solving nonlinear equations and apply it three ways to determine the maximum of a likelihood function.
Fitting GLMS

Three popular numerical methods are the following:

- Newton-Raphson method
- Fisher scoring method
- Reweighted Least Squares
Newton-Raphson method

- The **Newton-Raphson method** is an iterative method for solving nonlinear equations, such as equations whose solution determines the point at which a function takes its maximum.
- It begins with an initial guess for the solution.
- It obtains a second guess by approximating the function to be maximized in a neighborhood of the initial guess by a second-degree polynomial.
- Then we find the location of that polynomial’s maximum value.
- It then approximates the function in a neighborhood of the second guess by another second-order polynomial, and the third guess is the location of its maximum.
- In this manner, the method generates a sequence of guesses.
- These converge to the location of the maximum when the function is suitable and/or the initial guess is good.
Here’s how Newton-Raphson determines the value $\hat{\beta}$ at which a function $L(\beta)$ is maximized.

Let $u' = (\partial L(\beta)/\partial \beta_1, \partial L(\beta)/\partial \beta_2, \ldots)$

Let $H$ denote the matrix having entries $h_{ab} = \partial^2 L(\beta)/\partial \beta_a \partial \beta_b$, called the Hessian matrix.

Let $u^{(t)}$ and $H^{(t)}$ be $u$ and $H$ evaluated at $\beta^{(t)}$, the guess $t$ for $\hat{\beta}$.

Step $t$ in the iterative process ($t = 0, 1, 2, \ldots$) approximates $L(\beta)$ near $\beta^{(t)}$ by the terms up to second order in its Taylor series expansion

$$L(\beta) \approx L(\beta^{(t)}) + u^{(t)'}(\beta - \beta^{(t)}) + \left(\frac{1}{2}\right)(\beta - \beta^{(t)})' H^{(t)}(\beta - \beta^{(t)})$$

Solving $\partial L(\beta)/\partial \beta = u^{(t)} + H^{(t)}(\beta - \beta^{(t)}) = 0$ for $\beta$ yields the next guess.

That guess can be expressed as

$$\beta^{(t+1)} = \beta^{(t)} - (H^{(t)})^{-1} u^{(t)}$$

assuming that $H^{(t)}$ is nonsingular.
The convergence of $\beta^{(t)}$ to $\hat{\beta}$ for the Newton-Raphson method is usually fast.

For large $t$, the convergence satisfies, for each $j$,

$$|\beta_j^{(t+1)} - \hat{\beta}_j| \leq c|\beta_j^{(t)} - \hat{\beta}_j|^2$$

for some $c > 0$.

It is referred to as *second-order*.

This implies that the number of correct decimals in the approximation roughly doubles after sufficiently many iterations.
Newton-Raphson method for logistic regression

\[
L(\pi) = y \log(\pi) + (n - y) \log(1 - \pi)
\]

\[
u = \frac{(y - n\pi)}{\pi(1 - \pi)}
\]

\[
H = -\left[\frac{y}{\pi^2} + \frac{(n - y)}{(1 - \pi)^2}\right]
\]

\[
\pi^{(t+1)} = \pi^{(t)} + \left[\frac{y}{\pi^{(t)}_2} + \frac{n - y}{(1 - \pi^{(t)})^2}\right]^{-1}\left(\frac{y - n\pi^{(t)}}{\pi^{(t)}}\right)(1 - \pi^{(t)})
\]
Fisher Scoring Method

- Fisher scoring is an alternative iterative method for solving likelihood equations.
- It resembles the NR method, the distinction being with the Hessian matrix.
- Fisher scoring uses the expected value of this matrix, called the expected information, whereas NR uses the matrix itself, called observed information.
Fisher Scoring Method

Let $\mathcal{J}(t)$ denote the approximation $t$ for the ML estimate of the expected information matrix;

That is, $\mathcal{J}(t)$ has elements $-E(\partial^2 L(\beta)/\partial \beta_a \partial \beta_b)$, evaluated at $\beta(t)$.

The formula for Fisher scoring is

$$\beta^{(t+1)} = \beta^{(t)} + (\mathcal{J}(t))^{-1} u^{(t)}$$

or

$$\mathcal{J}(t) \beta^{(t+1)} = \mathcal{J}(t) \beta^{(t)} + u^{(t)}$$

NR and Fisher scoring algorithms are identical for canonical link models.
Fisher Scoring Method in logistic regression

- Information is $n/[\pi(1 - \pi)]$
- A step of Fisher scoring gives

\[
\pi^{(t+1)} = \pi^{(t)} + \left[ \frac{n}{\pi^{(t)}(1 - \pi^{(t)})} \right]^{-1} \frac{y - n\pi^{(t)}}{\pi^{(t)}(1 - \pi^{(t)})}
\]

\[
= \pi^{(t)} + \frac{y - n\pi^{(t)}}{n} = \frac{y}{n}
\]
Fisher Scoring Method in logistic regression

- \( \mathcal{J}^{(t)} = X' W^{(t)} X \), where \( W^{(t)} \) is \( W \) evaluated at \( \beta^{(t)} \)
- The estimated asymptotic covariance matrix \( \hat{J}^{-1} \) of \( \hat{\beta} \) occurs as a by-product of this algorithm as \( (\mathcal{J}^{(t)})^{-1} \) for \( t \) at which convergence is adequate.
- For both Fisher scoring and Newton-Raphson, \( u \) has elements

\[
u_j = \frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}\]
A relation exists between weighted least squares estimation and using Fisher scoring to find ML estimates.

We refer here to the general linear model of form

\[ z = X\beta + \epsilon \]

When the covariance matrix of \( \epsilon \) is \( V \), the weighted least squares (WLS) estimator of \( \beta \) is

\( (X'V^{-1}X)^{-1}X'V^{-1}z \)

From \( \mathbf{J} = X'WX \), expression for elements of \( u \), and since diagonal elements of \( W \) are

\[ w_i = \frac{(\partial \mu_i/\partial \eta_i)^2}{\text{Var}(Y_i)} \]

\[ \mathbf{J}(t)\beta(t) + u(t) = X'W(t)z(t) \]

where \( z(t) \) has elements

\[ z_i(t) = \sum_j x_{ij}\beta_j(t) + (y_i - \mu_i(t))\frac{\partial \eta_i(t)}{\partial \mu_i(t)} \]

\[ = \eta_i(t) + (y_i - \mu_i(t))\frac{\partial \eta_i(t)}{\partial \mu_i(t)} \]
ML as Iterative Reweighted Least Squares

Equations for Fisher scoring then have the form

\[(X'W(t)X)\beta^{(t+1)} = X'W(t)z(t)\]

These are the normal equations for using weighted least squares to fit a linear model for a response variable \(z^{(t)}\), when the model matrix is \(X\) and the inverse of the covariance matrix is \(W^{(t)}\).

The equations have solution

\[\beta^{(t+1)} = (X'W(t)X)^{-1}X'W(t)z(t)\]
ML as Iterative Reweighted Least Squares

- The vector $z$ in this formulation is a linearized form of the link function $g$ evaluated at $y$

  $$g(y_i) \approx g(\mu_i) + (y_i - \mu_i)g'(\mu_i) = \eta_i + (y_i - \mu_i)(\partial \eta_i / \partial \mu_i) = z_i$$

- This adjusted response variable $z$ has element $i$ approximated by $z_i^{(t)}$ for cycle $t$ of the iterative scheme.

- That cycle regresses $z^{(t)}$ on $X$ with weight (i.e., inverse covariance) $W^{(t)}$ to obtain a new estimate $\beta^{(t+1)}$.

- This estimate yields a new linear predictor value $\eta^{(t+1)} = X\beta^{(t+1)}$ and a new adjusted response value $z^{(t+1)}$ for the next cycle.

- The ML estimator results from iterative use of weighted least squares, in which the weight matrix changes at each cycle.

- The process is called **iterative re-weighted least squares**.