Firth Penalized Regression

Emmanuel Ogundimu

General description: Binary data are prevalent in numerous practical problems across various disciplines, notably in social and medical sciences. Generalized Linear Models (GLMs) are commonly employed to analyze such data, with logistic regression and probit regression being widely utilized techniques for modeling binary outcomes. However, accurate parameter estimation, particularly via Maximum Likelihood Estimation (MLE), can be challenging in finite samples due to biases arising from factors such as measurement errors, class imbalance, and separation phenomena. To address the issues of class imbalance and separation, an implicit approach proposed by David Firth (David Firth is a semi-retired professor of Statistics at Warwick) in 1993 adjusts the MLE procedures to reduce bias effectively. While primarily applied in the context of logistic regression, this research aims to extend the utility of the method to various statistical models, including continuous and binary sample selection models, censored regression, and random effect models.

The key idea of the Firth method is to add a penalization term to the log-likelihood function, which reduces the bias in MLE without substantially affecting the estimates' efficiency.

Suppose we have n independent observations $(\mathbf{x}_i, y_i) \in \mathbb{R}^p \times \{0,1\}$, $i = 1, \ldots, n$, which are generated from

$$y_i|\mathbf{x}_i \sim \text{Bernoulli}(F(\boldsymbol{\beta}^T\mathbf{x}_i)),$$

where $\mathbf{x}_i$ are the covariates, $F : \mathbb{R} \to (0,1)$ is a known link function, and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown vector of regression coefficients. Various models can be described depending on the link function, F. The maximum likelihood method estimates the parameters by maximizing the log-likelihood function associated with the regression model.

Consider the logistic regression model, the log-likelihood function for n subjects is given by:

$$\ell(\boldsymbol{\beta}) = \sum_{i=i}^{n} \Big[y_i\boldsymbol{\beta}^T\mathbf{x_i} - \log\Big(1 + \exp(\boldsymbol{\beta}^T\mathbf{x_i})\Big)\Big].$$

The Firth method modifies the log-likelihood function by introducing a penalty term based on the Jeffreys invariant prior:

$$\ell^*(\beta) = \ell(\beta) + \frac{1}{2}\log|I(\beta)|,$$

where $I(\beta)$ is the information matrix evaluated at β.

The method has also been used to solve the problem of separation in binary data. Separation occurs when one or more model covariates perfectly predict the outcome of interest.

Figure 1 shows plots of the fitted probabilities from standard and Firth logistic regression models using a dataset with separation. Firth's method demonstrated better predictions compared to the unrealistic Heaviside function produced by the logistic regression model.

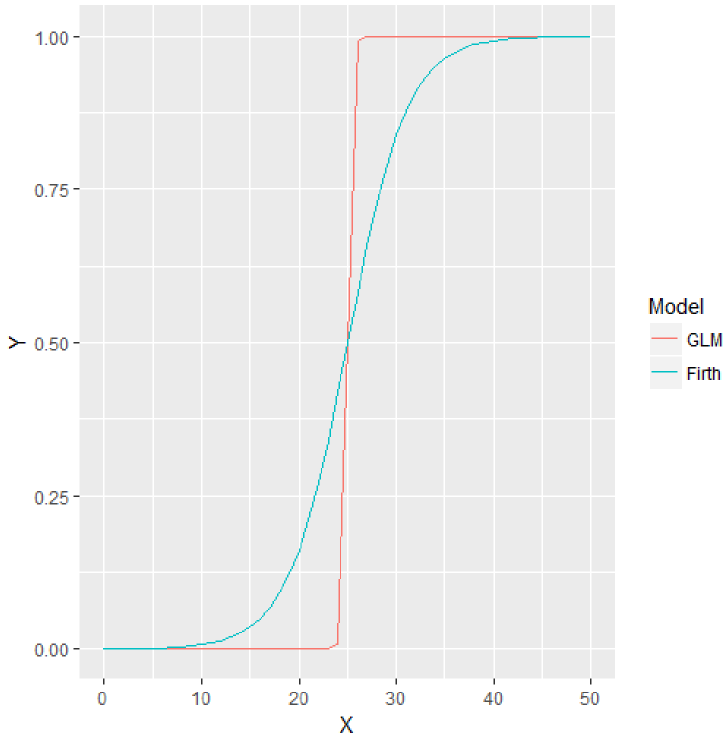

Figure 1: Predicted probabilities from logistic regression and Firth Method.

Objectives: This project aims to introduce students to the inner workings of the Firth method. We will explore the use of the method for various data types. Additionally, we will study the use of the model in low- and high-dimensional settings. Some potential directions for this project include:

- Exploration of various likelihood-based estimation techniques for Firth penalized regression such as the EM-algorithm, Newton-Raphson method, Trust Region, etc.
- Extension of the Firth method to various data types such as Tobit 1 and Tobit 2 models.
- Imputation of missing binary and survival data in the presence of separation and monotone likelihood.
- Improved Causal Inference with the Firth method.

Essential modules/Skills: MATH3411: Advanced Statistical Modelling; Familiarity with the R statistical software.

References

1. Firth, D. (1993). Bias reduction of maximum likelihood estimates. Biometrika 80(1), 27–38.
2. Heinze, G. and M. Schemper (2002). A solution to the problem of separation in logistic regression. Statistics in medicine 21(16), 2409–2419.
3. Kosmidis, I. (2014). Bias in parametric estimation: reduction and useful side-effects. Wiley Interdisciplinary Reviews: Computational Statistics 6(3), 185–196.
4. Ogundimu, E. O. (2019). Prediction of default probability by using statistical models for rare events. J. R. Statist. Soc. A. 182(4), 1143–1162.