Separation and Convergence Issues in Logistic Regression

Cornell Statistical Consulting Unit

1 Introduction

In logistic regression models, if the sample size is small or if a predictor is strongly associated with one of the possible outcomes (say, 0 or 1), then the algorithm which computes coefficient estimates becomes unreliable. The resulting regression coefficients may be biased and will have extremely large standard errors. A similar problem occurs in contingency tables when the sample size is small or when too many cells in the table have low counts. This is known as "the phenomenon of separation".

2 Example

Consider a logistic regression model with a binary outcome y (equal to 0 or 1) and a single continuous predictor, x. The logistic regression model is

$$\ln\left(\frac{P(Y=1 \mid x)}{1-P(Y=1 \mid x)}\right)=\beta_0+\beta_1 x$$

For this simulated example, the true regression coefficient for x is $\beta_1 = 1$. We have $n = 60$ simulated observations. The data and the fitted logistic regression model are depicted in Figure 2.1. We can see that most of the $y = 1$ observations occur when x is greater than zero. This predictor is strongly associated with the binary outcome variable (the probability of $y = 1$ seems to increase as x increases).

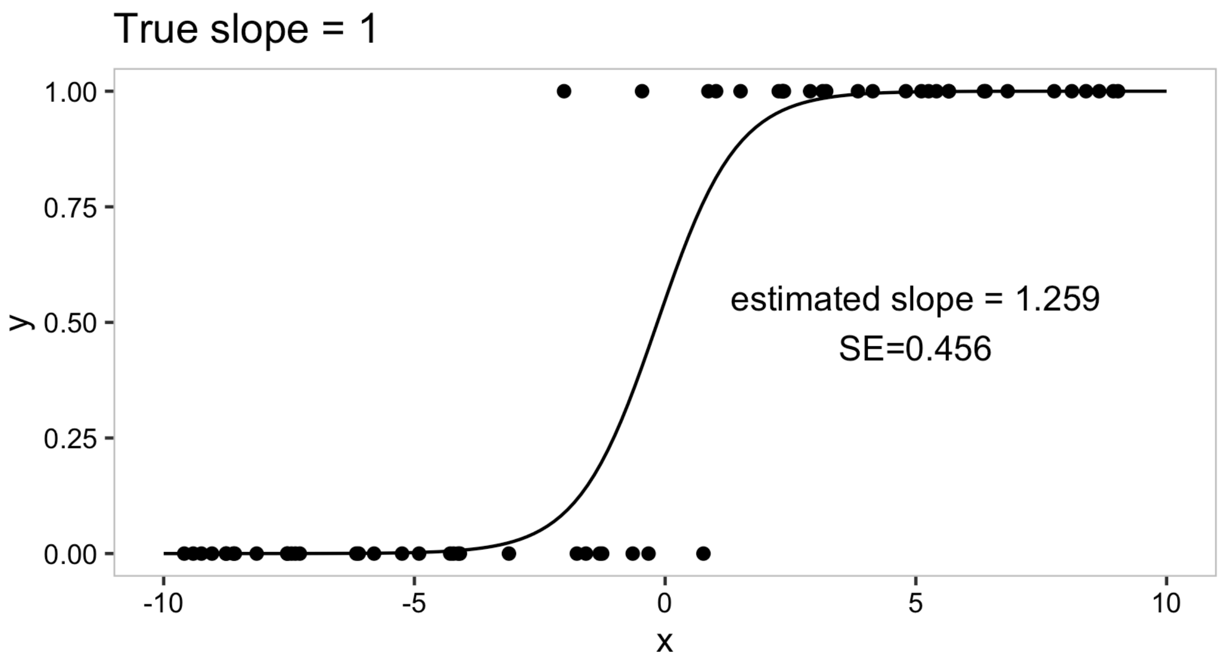

Figure 2.1: Fitted logistic regression model. The predictor is strongly associated with the response variable, but the data are not perfectly separated and there are no computational issues when computing coefficient estimates.

Consider what happens if we remove two of the $y = 1$ observations. Specifically, we will remove the two $y = 1$ observations with the smallest x values. When these observations are removed, then *all* of the $y = 1$ observations will have x values greater than $x = 0.76$. This is depicted in Figure 2.2. In this case, the predictor x "perfectly separates" the $y = 1$ observations from the $y = 0$ observations. When we try to fit the logistic regression model, the algorithm fails to converge and the coefficient estimate and standard error are extremely large.

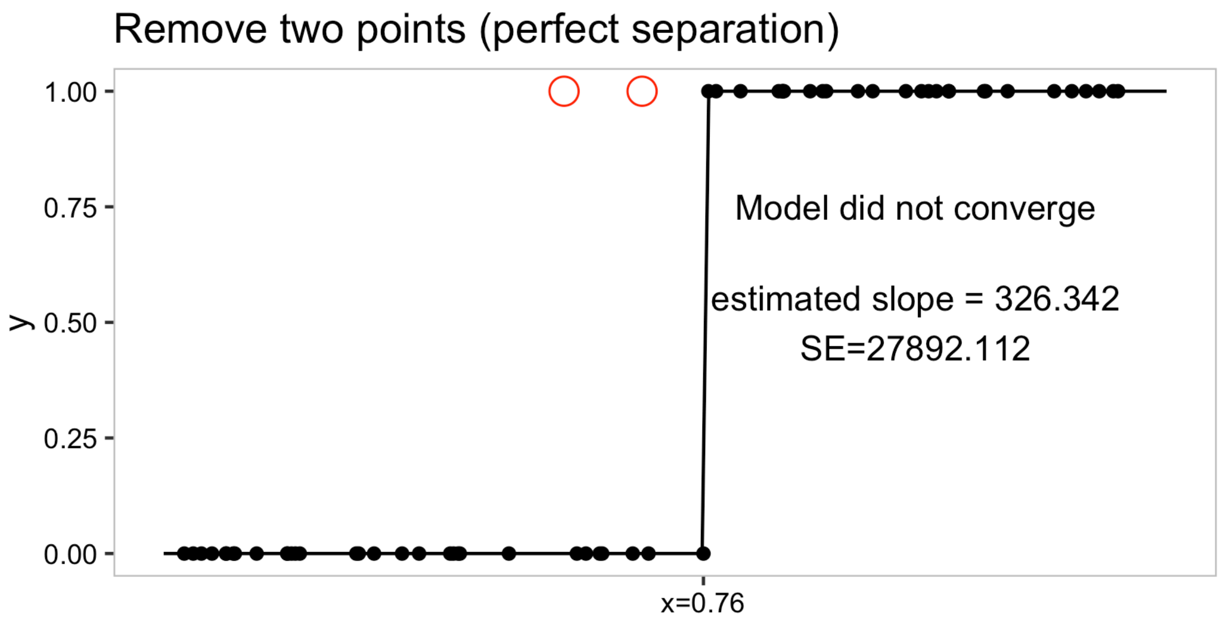

Figure 2.2: Logistic regression with separation. When two of the $y = 1$ observations are removed, then all of the $y = 1$ observations have x values greater than $x = 0.76$. So the observations are "perfectly separated" by the predictor variable and the logistic regression estimation algorithm will not converge.

3 Firth's biased-reduced logistic regression

One way to address the separation problem is to use Firth's bias-adjusted estimates (Firth 1993). In logistic regression, parameter estimates are typically obtained by maximum likelihood estimation. When the data are separated (or nearly so), the maximum likelihood estimates can be infinite and the algorithm will fail to converge. Firth's method maximizes a "penalized" likelihood function and does not suffer from the convergence issues of standard maximum likelihood in the presence of separation.

Figure 3.1 depicts the logistic regression model using Firth's method instead of standard maximum likelihood. When the data are perfectly separated after deleting two of the observations, Firth's method does not have any convergence issues and the standard error for the estimated slope is not inflated, as it was using standard maximum likelihood (Figure 2.2).

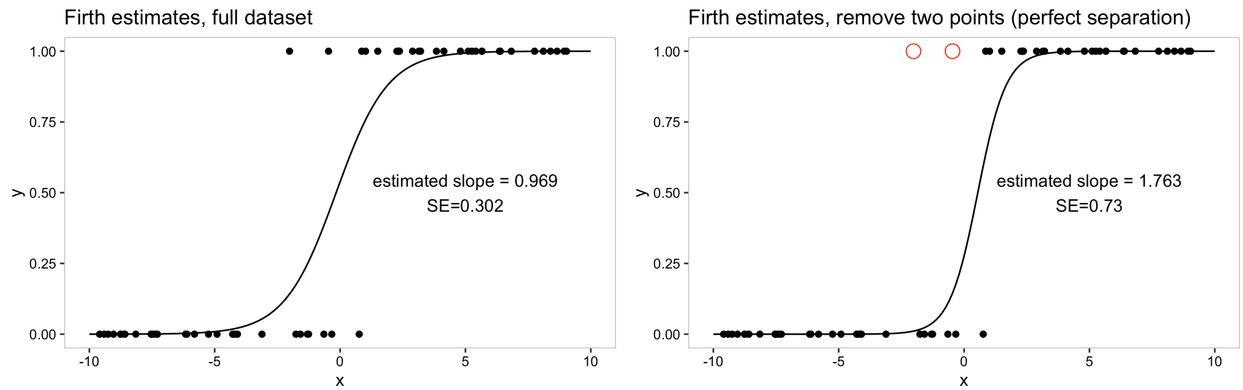

Figure 3.1: Fitting the logistic regression model usign Firth's method. Even with perfect separation (right panel), Firth's method has no convergence issues when computing coefficient estimates.

Firth's bias-adjusted estimates can be computed in JMP, SAS and R. In SAS, specify the FIRTH option in in the MODEL statement of PROC LOGISTIC. In JMP, these estimates are available in the *Fit Model* window: choose Generalized Linear Model for the model Personality, and check the box next to "Firth's Bias-Adjusted Estimates". In R, Firth's method is implemented in the *logistf* package.

Reference: Firth, D. (1993). "Bias reduction of maximum likelihood estimates", Biometrika 80, 27–38.

Created February 2012. Last updated April 2022.