Bias Reduction and a Solution for Separation of Logistic Regression with Missing Covariates

Tapabrata Maiti[1], **Vivek Pradhan**[2]

[1]Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA email: maiti@stt.msu.edu

[2]Cytel Inc, Cambridge, MA 02139, USA email: vpradhan@cytel.com

Abstract

Logistic regression is an important statistical procedure used in many disciplines. The standard software packages for data analysis are generally equipped with this procedure where the maximum likelihood estimates of the regression coefficients are obtained iteratively. It is well known that the estimates from the analyses of small or medium sized samples are biased. Also in finding such estimates, often a separation is encountered in which the likelihood converges but at least one of the parameter estimates diverge to infinity. Standard approaches of finding such estimates do not take care of these problems. Moreover, the missingness in the covariates adds an extra layer of complexity to the whole process. In this article we address these three practical issues - bias, separation, and missing covariates by means of simple adjustments. We have applied the proposed technique using real and simulated data. The proposed method always find a solution and the estimates are less biased. A SAS macro that implements the proposed method can be obtained from the authors.

Key Words: Small-sample, Bias, EM-algorithm, Maximum likelihood estimate