Comparison of robust estimators in the negative binomial model with an application to falls data

Falls occur frequently among certain categories of patients, e.g. elder people who have recently been hospitalized or people suffering from Parkinson’s disease. As these falls have devastating consequences, researchers in musculoskeletal diseases regularly conduct intervention studies aiming at reducing their occurrence. However, the data is often plagued by the so-called multiple fallers, i.e. patients that seem to fall a lot more than others, the reasons for this being unknown, and these high values in the response may seriously bias estimation and inference. The number of falls is typically modelled using negative binomial (NB) regression with classical estimation methods (maximum likelihood and moment-based estimators) well-known to be sensitive to outliers and model misspecifications. In the falls literature, the problem of multiple fallers is typically dealt with ad hoc procedures such as truncating the number of falls at 12. We examine here two robust alternatives for estimating the regression coefficients in the NB model. The first one directly follows Cantoni and Ronchetti (2001), where robustness is achieved by applying a bounded function on the Pearson residuals. A similar approach has been used for M-quantile regression with application to NB regression by Chambers et al. (2012). However, we propose a robust weighted maximum likelihood estimator for the over dispersion parameter and consider an iterative scheme between estimating the latter and the regression coefficients. Moreover, we inspect both Huber’s and Hampel’s three-part (redescending) -functions. The second alternative extends the work of Bianco et al. (2012), where the individual deviance components are bounded. In this work in progress, we will compare the performance of the different estimators and explain what has been learnt so far. An application to recent fall trial data will also be presented.