A Robust Approach for Multinomial Logistic Regression Model Application to Mammography Experience Data

Elena Castilla & Pedro J. Chocano

Complutense University of Madrid, Spain

UNIVERSIDAD COMPLUTENSE MADRID

The Multinomial Logistic Regression Model

Let us assume that we have a nominal outcome variable $\boldsymbol{Y}$ with $d+1$ categories $C_1, \ldots, C_{d+1}$ depedning on k explanatory variables with given values $x_1, \ldots, x_k$. Let $\pi_j(\boldsymbol{x};\boldsymbol{\beta})$ denote the probability that $\boldsymbol{Y}$ belongs to the category C_j for $j=1,\ldots,d+1$. The multinomial logistic regression model (MLRM) is given by

$$\pi_j(\boldsymbol{x};\boldsymbol{\beta}) = \Pr(Y_j = 1|\boldsymbol{x}) = \begin{cases} \dfrac{\exp\{\boldsymbol{x}^T\boldsymbol{\beta}_j\}}{1+\sum_{l=1}^{d}\exp\{\boldsymbol{x}^T\boldsymbol{\beta}_l\}}, & j=1,...,d \\ \dfrac{1}{1+\sum_{l=1}^{d}\exp\{\boldsymbol{x}^T\boldsymbol{\beta}_l\}}, & j=d+1 \end{cases},$$

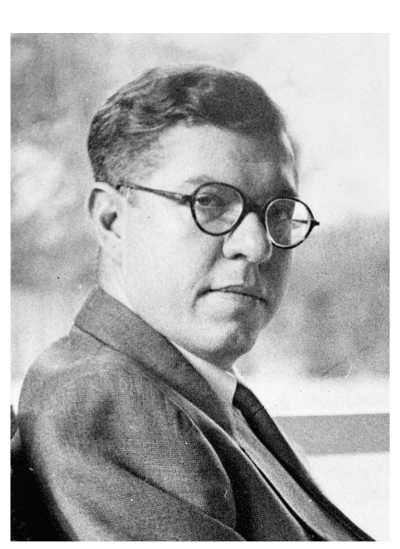
with $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, ..., \beta_{jk})^T \in \mathbb{R}^{k+1}$, $j=1,...,d$. Since the full parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_d^T)^T$ is $d(k+1)$-dimensional, the parameter space is $\Theta = \mathbb{R}^{d(k+1)}$. Here, the vector of explanatory variables takes the value $\boldsymbol{x} = (x_0, x_1, \ldots, x_k)^T$ with $x_0 = 1$ being associated with the intercept β_{j0} for $j=1,...,d$.

Now, assume that we have observed the data on N independent individuals with associate covariate values $\boldsymbol{x}_i \in \mathbb{R}^{k+1}$ having responses $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{i,d+1})^T$, for $i=1,\ldots,N$. The most common estimator of $\boldsymbol{\beta}$ for the MLRM is the maximum likelihood estimator (MLE), $\hat{\boldsymbol{\beta}}$, obtained by maximizing the loglikelihood function

$$\log \mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{N}\sum_{j=1}^{d+1} y_{ij}\log \pi_j(\boldsymbol{x}_i;\boldsymbol{\beta}).$$

The Robustness Problem

The main problem of the MLE is its lack of robustness, i.e., its poor behaviour in the presence of outliers, which are observations that lie an abnormal distance from other values. While a first temptation would be that of omitting these data points, it is not acceptable to drop an observation just because it is an outlier since we can loose important information. Thus, we need of robust approaches to model our data.



I don't see the logic of rejecting data just because they seem incredible.

Sir Fred Hoyle
1915–2001

Our approach

We define a new family of robust estimators and Wald-type tests based on the minimum Cressie-Read ϕ-divergence measures, depending on a tuning parameter λ. For $\lambda = 0$ we obtain the classical MLE and Wald-test as a special case.

The Mammography Experience Data (I)

The Mammography experience data, a subset of a study by the University of Massachusetts Medical School, assess factors associated with women's knowledge, attitude and behavior towards mammography. This study involves 412 individuals, grouped in 125 distinct covariates values and 8 explanatory variables. The response variable ME (Mammography experience) is a categorical factor with three levels: "Never", "Within a Year" and "Over a Year". The groups of observations associated with covariate values $\boldsymbol{x}_i$ for $i \in \{1,3,17,35,75,81,102\}$ can be treated as outliers.

We compute the minimum ϕ-divergence estimators of $\boldsymbol{\beta}$ for $\lambda = 0$ (MLE) and $\lambda = -0.5$. Moreover, we plot the corresponding (estimated) category probabilities for each available distinct covariate values. The left panel of Figure 1 presents these probabilities for the first category and $\lambda = 0$, while the right panel presents these category probabilities for $\lambda = -0.5$.

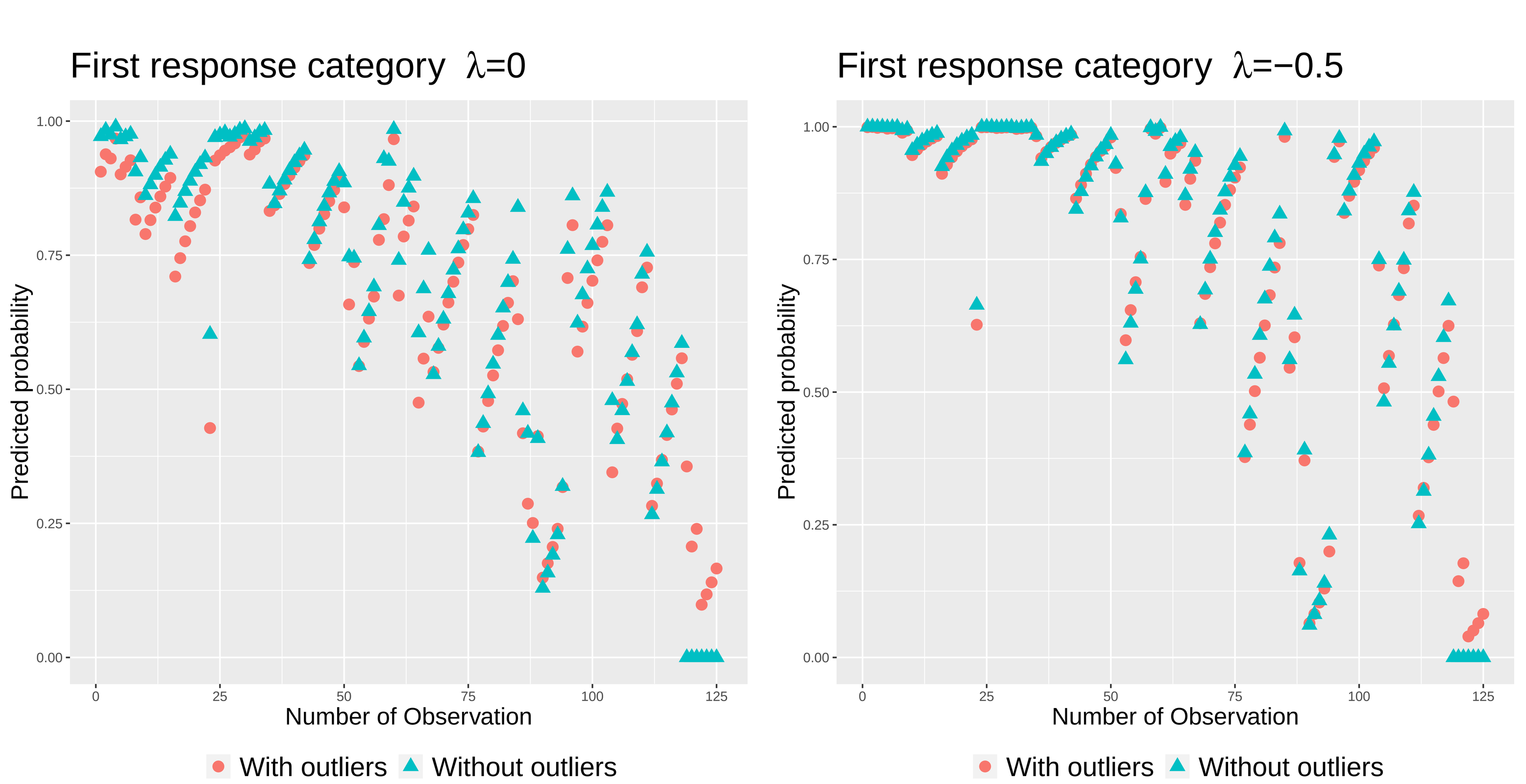


Figure 1: Predicted category probabilities of the response variable

The Mammography Experience Data (II)

Now, we want to evaluate the robustness of the proposed Wald-type tests. We consider the problem of testing

$$H_0: \beta_{SYMPT_{21}} = 0,$$
$$H_0: \beta_{SYMPT_{11}} = \beta_{SYMPT_{21}},$$

for the variable SYMPT$_{r,j}$ ("You do not need a mammogram unless you develop symptoms: $r=1$, strongly agree; $r=2$, agree; $r=3$, disagree; $r=4$, strongly disagree). The p-values obtained based on the proposed test are plotted over λ in Figure 2 for both the full and the outlier deleted data. We may reject the first hypothesis while the second hypothesis may be accepted.

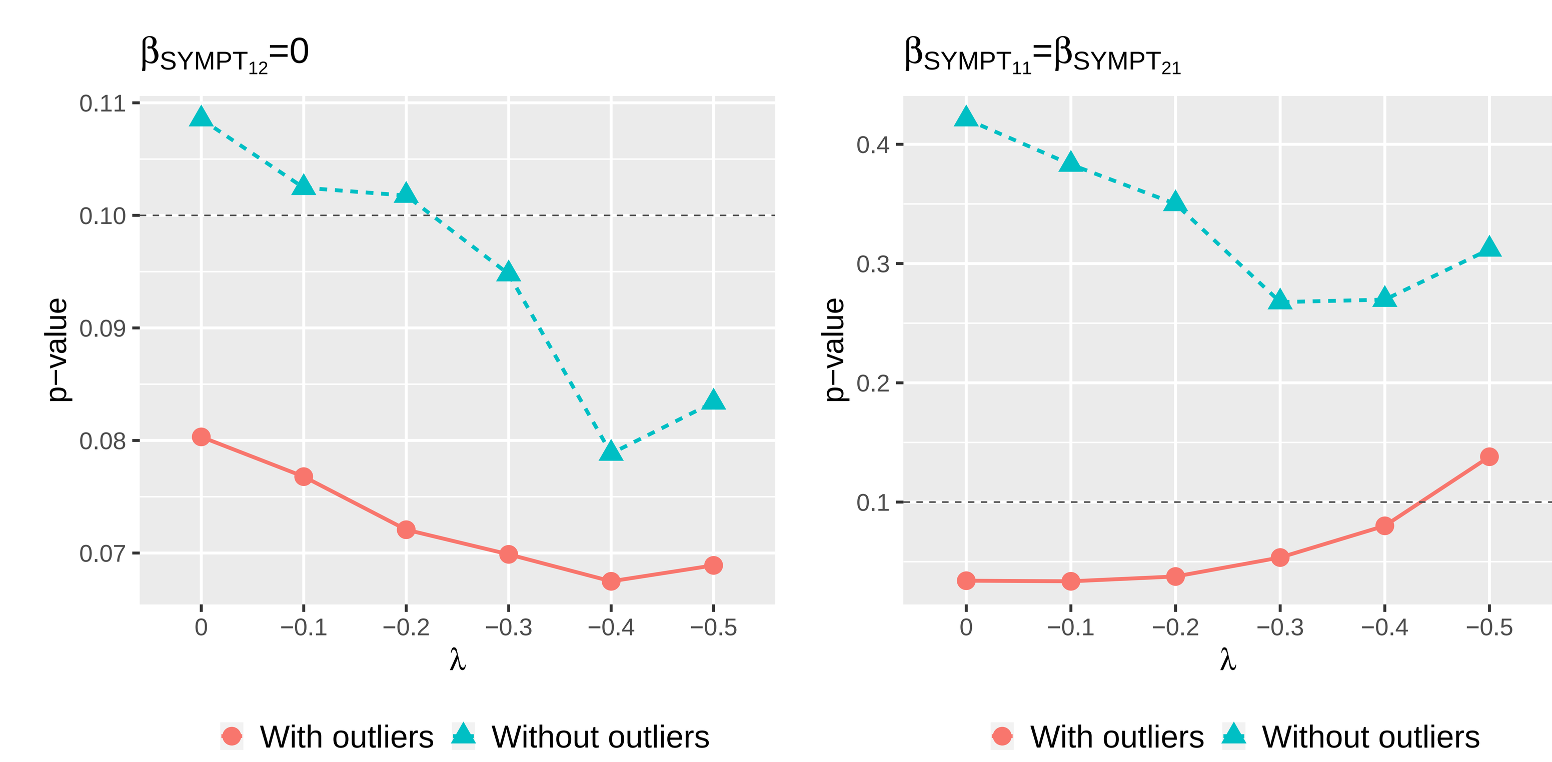


Figure 2: p-values of the proposed Wald-type tests.

Performance of the proposed approach

Results clearly indicate the significant variation of the MLE in the presence or absence of the outliers. However, the estimator with $\lambda = -0.5$ is shown to be much more stable. On the other hand, test decisions at the significance level $\alpha = 0.1$ change completely in the presence of outliers for λ near to 0.

Main conclusions

Attending to the results we can conclude that women who have never had a mammogram are very influenziated by the thought that you only need it in case of developing symptoms.