On the Asymptotic Distribution of Cook's distance in Logistic Regression Models*

Nirian Martín† and Leandro Pardo‡
Complutense University of Madrid

May 12, 2008

Abstract

It sometimes occurs that one or more components of the data exerts a disproportionate influence on the model estimation. We need a reliable tool for identifying such troublesome cases in order to decide either eliminate from the sample, when the data collect was badly realized, or otherwise take care on the use of the model because the results could be affected by such components. Since a measure for detecting influential cases in linear regression setting was proposed by Cook [7], apart from the same measure for other models, several new measures have been suggested as single-case diagnostics. For most of them some cutoff values have been recommended (see Belsley et al. [4], for instance), however the lack of a quantile type cutoff for Cook's statistics has induced the analyst to deal only with index plots as worthy diagnostics tools. Focussed on logistic regression, the aim of this paper is to provide the asymptotic distribution of Cook's distance in order to look for a meaningful cutoff point for detecting influential and leverage observations.

1 Introduction

Logistic regression is a model associated with I integer responses y_i, $i = 1, ..., I$, each of them the number of successful observations of n_i trials, which come from independent Binomial random variables Y_i, $i = 1, ..., I$. This model establishes that the probabilities of success $\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}) = \Pr(Y_i = 1)$, $i = 1, ..., I$, depend on unknown parameters $\boldsymbol{\beta} = (\beta_0, ..., \beta_k)^T$ and as well as explanatory variables $\boldsymbol{x}_i = (x_{i0}, ..., x_{ik})^T$, $x_{i0} = 1$, $i = 1, ..., I$, according to the formula

$$\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}) = \frac{\exp\{\boldsymbol{x}_i^T\boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_i^T\boldsymbol{\beta}\}}, \quad i = 1, ..., I. \tag{1}$$

We denote by $\boldsymbol{X}$ the $I \times (k+1)$ matrix with rows $\boldsymbol{x}_i^T$, $i = 1, ..., I$. This matrix involve I realizations, $(x_{1h}, ..., x_{Ih})^T \in \mathbb{R}^I$, for k real explanatory variables X_h, $h = 2, ..., k+1$, i.e. once an intercept β_0 is included by setting a first column of $\boldsymbol{X}$ equals $(1, ..., 1)^T$, the rest of the columns of $\boldsymbol{X}$ correspond with a sample of X_h, $h = 2, ..., k+1$, and all samples belong to a common set of I individuals. We also shall assume that $\operatorname{rank}(\boldsymbol{X}) = k + 1$.

The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ is defined by

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} \log \mathcal{L}(\boldsymbol{\beta}), \tag{2}$$

*This work was partially supported by Grants MTM2006-06872 and CAM-UCM2007-910707.
†School of Statistics, E-mail: nirian@estad.ucm.es (Corresponding author).
‡Mathematics Faculty, E-mail: lpardo@mat.ucm.es.

where $\Theta = \{\boldsymbol{\beta}= (\beta_0, ..., \beta_k)^T : \beta_i \in (-\infty, \infty), \ i = 0, ..., k\}$ is the parameter space and $\mathcal{L}(\boldsymbol{\beta})$ is the likelihood function for the logistic regression model

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{I} \binom{n_i}{y_i} \pi(\boldsymbol{x}_i^T \boldsymbol{\beta})^{y_i} (1 - \pi(\boldsymbol{x}_i^T \boldsymbol{\beta}))^{n_i - y_i}.$$

It is well-known that $\widehat{\boldsymbol{\beta}}$ can be obtained as the solution $\boldsymbol{\beta}$ of the nonlinear system of equations

$$\boldsymbol{X}^T (\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{0}_{(k+1)\times 1}, \tag{3}$$

being $\boldsymbol{y} = (y_1, ..., y_I)^T$ and $\boldsymbol{\mu} = (n_1 \pi (\boldsymbol{x}_1^T \boldsymbol{\beta}), ..., n_I \pi (\boldsymbol{x}_I^T \boldsymbol{\beta}))^T$. For more details about logistic regression model see Agresti [1] and references therein.

Let $N = \sum_{i=1}^{I} n_i$ be the sum of trials considered in the model. If we omit all the trials associated with the random variable Y_j ($j \in \{1, ..., I\}$), we have in total $N - n_j$ trials and the MLE of $\boldsymbol{\beta}$ is defined by

$$\widehat{\boldsymbol{\beta}}^{(j)} = \arg \min_{\boldsymbol{\beta} \in \boldsymbol{\Theta}} \log \mathcal{L}^{(j)}(\boldsymbol{\beta}), \tag{4}$$

where

$$\mathcal{L}^{(j)}(\boldsymbol{\beta}) = \prod_{\substack{i=1 \\ i \neq j}}^{I} \binom{n_i}{y_i} \pi(\boldsymbol{x}_i^T \boldsymbol{\beta})^{y_i} (1 - \pi(\boldsymbol{x}_i^T \boldsymbol{\beta}))^{n_i - y_i}.$$

It is well-known that influential observations in logistic regression models are those points which greatly change the results of the statistical analysis when omitted from the sample. Johnson [16] considered the following influence measure associated with the j-th observation of the logistic regression model

$$D^{(j)}(\widehat{\boldsymbol{\beta}}) = N(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{(j)})^T \boldsymbol{X}^T \boldsymbol{W}_N(\widehat{\boldsymbol{\beta}}) \boldsymbol{X} (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{(j)}), \quad j \in \{1, ..., I\}, \tag{5}$$

where

$$\boldsymbol{W}_N(\boldsymbol{\beta}) = \operatorname{diag}\left(\left(\tfrac{n_i}{N} \pi(\boldsymbol{x}_i^T \boldsymbol{\beta}) \left(1 - \pi(\boldsymbol{x}_i^T \boldsymbol{\beta})\right)\right)_{i=1,...,I} \right). \tag{6}$$

The influence measure $D^{(j)}(\widehat{\boldsymbol{\beta}})$ is the natural adaptation, to the context of logistic regression, of Cook's distance (see Cook [7]) for detecting influential observations in linear regression. Based on the similarity of $D^{(j)}(\widehat{\boldsymbol{\beta}})$ with the quadratic form,

$$N(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}_0) \boldsymbol{X} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

where $\boldsymbol{\beta}_0$ is the true value of the vector of unknown parameters in the logistic regression model and $\boldsymbol{W}(\boldsymbol{\beta}_0) = \lim_{N\to\infty} \boldsymbol{W}_N(\boldsymbol{\beta}_0)$, Johnson considered that the asymptotic distribution of $D^{(j)}(\widehat{\boldsymbol{\beta}})$ can be compared with a chi-square distribution with $k+1$ degrees of freedom. This idea of Johnson was based on some ideas considered by some authors in linear regression. In linear regression models with N observable normal distributed responses and k unknown parameters for a full rank design matrix, Cook [7] and Weisberg [23, page 108] recommended the 50-th percentile of the F distribution with (k,$N-k$) degrees of freedom as a rule-of-thumb for determining influential observations. This rule has been used for many years until on one hand Muller and Chen Mok [17] and on the other hand Jensen and Ramirez [15] presented the statistical theory for using Cook's distance as a test for determining influential observations by deriving the exact distribution of Cook's distance. While the first authors considered the classical Cook's distance, the others took into account the version related with Welsch-Kuh's distance (DFFITS). Following the same idea to distinguish the internally and externally Studentized residuals, the essential difference between both of them is based on the sample to build the estimator of the variance, i.e. it is considered the fact of maintaining or deleting the observation studied by Cook's distance. The exact distribution of the classical Cook's distance multiplied by a constant is the F-Snedecor distribution with

(1,$N-k-1$) degrees of freedom and the square of the Welsch-Kuh's distance multiplied by a constant is the Beta distribution with (1/2,$(N-k-1)/2$) degrees of freedom. Furthermore, Muller and Chen Mok [17] considered random coefficients rather than fixed coefficients $x_{i0},...,x_{ik}$ $(i\in\{1,...,I\})$. Recently these results are being extending to different models as for example generalizations of the linear regression (see Díaz-García et al. [9]).

Pregibon [21] considered the one-step Newton-Raphson method for logistic regression with initial point $\widehat{\boldsymbol{\beta}}$ in order to get an approximation, $\widehat{\boldsymbol{\beta}}_*^{(j)}$, of $\widehat{\boldsymbol{\beta}}^{(j)}$. Based on $\widehat{\boldsymbol{\beta}}_*^{(j)}$ he got an approximation, $\widetilde{D}^{(j)}(\widehat{\boldsymbol{\beta}})$, of $D^{(j)}(\widehat{\boldsymbol{\beta}})$ and even it seems that it performs accurately in most cases he demonstrated through an example that however this approximation can be not very good (see Figure 7 in [21] which is referred to Example 1 in Section 5 of this paper). Remind that in linear regression there exists a version of the same formula in which an equality holds rather than an approximation (see for instance, Rao and Toutenburg [22, Section 7.5]).

Our interest in this paper is focussed on obtaining the asymptotic distribution of $D^{(j)}(\widehat{\boldsymbol{\beta}})$ through the definition of $\widehat{\boldsymbol{\beta}}^{(j)}$ given in (4). Section 2 is devoted to provide the expressions of some asymptotic results that will be necessary in Section 3 which is the most important part of the paper because in it we present the main result from which arises the idea of decompose the distribution of Cook's distance in interpretable components. In Section 4 a distribution based cutoff is presented and Section 5 is devoted to apply such cutoff values for some well-known examples in which some troublesome observations are encountered. Furthermore, the same examples will be useful to investigate more thoroughly the appropriateness of approximating $D^{(j)}(\widehat{\boldsymbol{\beta}})$ by $\widetilde{D}^{(j)}(\widehat{\boldsymbol{\beta}})$.

2 Some distributional results in logistic regression models

Let

$$\widehat{\boldsymbol{p}}_N \equiv \left(\frac{y_1}{N},\frac{n_1-y_1}{N},\frac{y_2}{N},\frac{n_2-y_2}{N},....,\frac{y_I}{N},\frac{n_I-y_I}{N}\right)^T,$$

$$\boldsymbol{p}_N(\boldsymbol{\beta}_0) \equiv \left(\tfrac{n_1}{N}\pi(\boldsymbol{x}_1^T\boldsymbol{\beta}_0),\tfrac{n_1}{N}(1-\pi(\boldsymbol{x}_1^T\boldsymbol{\beta}_0)),...,\tfrac{n_I}{N}\pi(\boldsymbol{x}_I^T\boldsymbol{\beta}_0),\tfrac{n_I}{N}(1-\pi(\boldsymbol{x}_I^T\boldsymbol{\beta}_0))\right)^T,$$

two probability vectors, the first one composed by normalized relative frequencies and the second one by normalized theoretical probabilities. It is well-known (see for instance Pardo et al. [20, Section 2]) that

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \boldsymbol{\beta}_0 + \left(\boldsymbol{X}^T\boldsymbol{W}_N(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\bigoplus_{i=1}^{I}\boldsymbol{c}_{N,i}(\boldsymbol{\beta}_0)^T\operatorname{diag}\left(\boldsymbol{p}_N(\boldsymbol{\beta}_0)^{-1/2}\right)(\widehat{\boldsymbol{p}}_N-\boldsymbol{p}_N(\boldsymbol{\beta}_0))+\mathrm{o}(\|\widehat{\boldsymbol{p}}_N-\boldsymbol{p}_N(\boldsymbol{\beta}_0)\|)\\ &= \boldsymbol{\beta}_0 + \left(\boldsymbol{X}^T\boldsymbol{W}_N(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\bigoplus_{i=1}^{I}\boldsymbol{c}_i^*(\boldsymbol{\beta}_0)^T(\widehat{\boldsymbol{p}}_N-\boldsymbol{p}_N(\boldsymbol{\beta}_0))+\mathrm{o}(\|\widehat{\boldsymbol{p}}_N-\boldsymbol{p}_N(\boldsymbol{\beta}_0)\|), \qquad (7)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{c}_{N,i}(\boldsymbol{\beta}_0) &\equiv \left(\tfrac{n_i}{N}\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\left(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\right)\right)^{1/2}\begin{pmatrix}\left(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\right)^{1/2}\\ -\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)^{1/2}\end{pmatrix},\ i=1,....,I,\\ \boldsymbol{c}_i^*(\boldsymbol{\beta}_0) &\equiv \begin{pmatrix}(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0))\\ -\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\end{pmatrix},\ i=1,....,I, \qquad (8)\end{aligned}$$

and $\boldsymbol{W}_N(\boldsymbol{\beta}_0)$ was defined in (6). Based on Limit Central Theorem we have

$$\sqrt{N}\left(\widehat{\boldsymbol{p}}_N-\boldsymbol{p}_N(\boldsymbol{\beta}_0)\right)\xrightarrow[N\to\infty]{\mathcal{L}}\mathcal{N}\left(\boldsymbol{0}_{2I\times 1},\boldsymbol{\Sigma}\right), \qquad (9)$$

where

$$\boldsymbol{\Sigma} \equiv \bigoplus_{i=1}^{I} \left((\lambda_i \boldsymbol{\Sigma}_{\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)})_{i=1,...,I} \right), \tag{10}$$

$$\boldsymbol{\Sigma}_{\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)} \equiv \begin{pmatrix} \pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\left(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\right) & -\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\left(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\right) \\ -\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\left(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\right) & \pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\left(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\right) \end{pmatrix} \tag{11}$$

and

$$\lambda_i \equiv \lim_{N\to\infty} \frac{n_i}{N}.$$

By using (7) and (9), it is not difficult to establish that

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \underset{N\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(\mathbf{0}_{(k+1)\times 1}, (\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X})^{-1}\right) \tag{12}$$

where

$$\boldsymbol{W}(\boldsymbol{\beta}_0) = \lim_{N\to\infty} \boldsymbol{W}_N(\boldsymbol{\beta}_0) = \operatorname{diag}\left(\left(\lambda_i \pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\left(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\right)\right)_{i=1,...,I} \right).$$

We assume now that all the trials associated with the random variable Y_j ($j \in \{1, ..., I\}$), n_j, have been deleted. In this setting we denote the normalized vector or frequencies and the normalized vector of theoretical probabilities by

$$\widehat{\boldsymbol{p}}_N^{(j)} \equiv \left(\frac{y_1}{N-n_j}, \frac{n_1-y_1}{N-n_j}, ..., \frac{y_{j-1}}{N-n_j}, \frac{n_{j-1}-y_{j-1}}{N-n_j}, \frac{y_{j+1}}{N-n_j}, \frac{n_{j+1}-y_{j+1}}{N-n_j}, ..., \frac{y_I}{N-n_j}, \frac{n_I-y_I}{N-n_j} \right)^T,$$

$$\begin{aligned} \boldsymbol{p}_N^{(j)}(\boldsymbol{\beta}_0) \equiv \Big(& \tfrac{n_1}{N-n_j}\pi(\boldsymbol{x}_1^T\boldsymbol{\beta}_0), \tfrac{n_1}{N-n_j}(1-\pi(\boldsymbol{x}_1^T\boldsymbol{\beta}_0)), ..., \tfrac{n_{j-1}}{N-n_j}\pi(\boldsymbol{x}_{j-1}^T\boldsymbol{\beta}_0), \tfrac{n_{j-1}}{N-n_j}(1-\pi(\boldsymbol{x}_{j-1}^T\boldsymbol{\beta}_0)), \\ & \tfrac{n_{j+1}}{N-n_j}\pi(\boldsymbol{x}_{j+1}^T\boldsymbol{\beta}_0), \tfrac{n_{j+1}}{N-n_j}(1-\pi(\boldsymbol{x}_{j+1}^T\boldsymbol{\beta}_0)), ..., \tfrac{n_I}{N-n_j}\pi(\boldsymbol{x}_I^T\boldsymbol{\beta}_0), \tfrac{n_I}{N-n_j}(1-\pi(\boldsymbol{x}_I^T\boldsymbol{\beta}_0)) \Big)^T, \end{aligned}$$

and we have

$$\widehat{\boldsymbol{\beta}}^{(j)} = \boldsymbol{\beta}_0 + \left(\boldsymbol{X}_{(j)}^T \boldsymbol{W}_N^{(j)}(\boldsymbol{\beta}_0) \boldsymbol{X}_{(j)} \right)^{-1} \boldsymbol{X}_{(j)}^T \bigoplus_{\substack{i=1 \\ i\neq j}}^{I} \boldsymbol{c}_i^*(\boldsymbol{\beta}_0)^T \left(\widehat{\boldsymbol{p}}_N^{(j)} - \boldsymbol{p}_N^{(j)}(\boldsymbol{\beta}_0) \right) + \mathrm{o}\left(\left\| \widehat{\boldsymbol{p}}_N^{(j)} - \boldsymbol{p}_N^{(j)}(\boldsymbol{\beta}_0) \right\| \right),$$

where $\boldsymbol{c}_i^*(\boldsymbol{\beta}_0)$ is given by (8), $\boldsymbol{X}_{(j)}$ is obtained through $\boldsymbol{X}$ by deleting the j-th row and

$$\boldsymbol{W}_N^{(j)}(\boldsymbol{\beta}_0) \equiv \operatorname{diag}\left(\left(\tfrac{ni}{N-n_i} \pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\left(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)\right) \right)_{\substack{i=1,...,I \\ i\neq j}} \right).$$

3 The distribution of Cook's distance

The next theorem establishes the asymptotic distribution of the difference between the estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^{(j)}$.

Theorem 1 *Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^{(j)}$ the MLE of parameters $\boldsymbol{\beta}$ based on the full observations and the MLE based on the full observations minus j-th observation. Then*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{(j)}) \underset{N\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(\mathbf{0}_{(k+1)\times 1}, \boldsymbol{\Sigma}^{(j)}\right) \tag{13}$$

where

$$\boldsymbol{\Sigma}^{(j)} \equiv \tfrac{w_j(\boldsymbol{\beta}_0)}{1-h_{jj}(\boldsymbol{\beta}_0)} \left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X} \right)^{-1} \boldsymbol{x}_j \boldsymbol{x}_j^T \left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X} \right)^{-1}$$

being $h_{jj}(\boldsymbol{\beta}_0) \equiv w_j(\boldsymbol{\beta}_0)^{1/2}\boldsymbol{x}_j^T \left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X} \right)^{-1} \boldsymbol{x}_j w_j(\boldsymbol{\beta}_0)^{1/2}$ and $w_j(\boldsymbol{\beta}_0) \equiv \lambda_j \pi(\boldsymbol{x}_j^T\boldsymbol{\beta}_0)\left(1-\pi(\boldsymbol{x}_j^T\boldsymbol{\beta}_0)\right)$ is the j-th diagonal element of $\boldsymbol{W}(\boldsymbol{\beta}_0)$, $j \in \{1, ..., I\}$.

Proof. Denoting each component of (8) by $c_{j1}^{*}(\boldsymbol{\beta}_0)$ and $c_{j2}^{*}(\boldsymbol{\beta}_0)$ respectively, and the j-th unit vector in $\mathbb{R}^{2I}$ by $\boldsymbol{e}_j = (0,...,0,1^{j)},0,...,0)^T$, we have

$$\boldsymbol{X}_{(j)}^T \bigoplus_{\substack{i=1\\ i\neq j}}^{I} \boldsymbol{c}_i^{*}(\boldsymbol{\beta}_0)^T \widehat{\boldsymbol{p}}_N^{(j)} = \boldsymbol{X}^T \left(\bigoplus_{i=1}^{I} \boldsymbol{c}_i^{*}(\boldsymbol{\beta}_0)^T - \boldsymbol{e}_j\, \boldsymbol{e}_{2j-1}^T c_{j1}^{*}(\boldsymbol{\beta}_0) - \boldsymbol{e}_j \boldsymbol{e}_{2j}^T c_{j2}^{*}(\boldsymbol{\beta}_0) \right) \tfrac{N}{N-n_j} \widehat{\boldsymbol{p}}_N$$

$$= \boldsymbol{X}^T \left(\bigoplus_{i=1}^{I} \boldsymbol{c}_i^{*}(\boldsymbol{\beta}_0)^T - \boldsymbol{e}_j \left(\boldsymbol{e}_{2j-1}^T c_{j1}^{*}(\boldsymbol{\beta}_0) + \boldsymbol{e}_{2j}^T c_{j2}^{*}(\boldsymbol{\beta}_0) \right) \right) \tfrac{N}{N-n_j} \widehat{\boldsymbol{p}}_N$$

and

$$\boldsymbol{X}_{(j)}^T \bigoplus_{\substack{i=1\\ i\neq j}}^{I} \boldsymbol{c}_i^{*}(\boldsymbol{\beta}_0)^T \boldsymbol{p}_N^{(j)}(\boldsymbol{\beta}_0) = \boldsymbol{X}^T \left(\bigoplus_{i=1}^{I} \boldsymbol{c}_i^{*}(\boldsymbol{\beta}_0)^T - \boldsymbol{e}_j \left(\boldsymbol{e}_{2j-1}^T c_{j1}^{*}(\boldsymbol{\beta}_0) + \boldsymbol{e}_{2j}^T c_{j2}^{*}(\boldsymbol{\beta}_0) \right) \right) \tfrac{N}{N-n_j} \boldsymbol{p}_N(\boldsymbol{\beta}_0).$$

Therefore,

$$\boldsymbol{X}_{(j)}^T \bigoplus_{\substack{i=1\\ i\neq j}}^{I} \boldsymbol{c}_i^{*}(\boldsymbol{\beta}_0)^T \left(\widehat{\boldsymbol{p}}_N^{(j)} - \boldsymbol{p}_N^{(j)}(\boldsymbol{\beta}_0) \right) = \boldsymbol{X}^T \left(\bigoplus_{i=1}^{I} \boldsymbol{c}_i^{*}(\boldsymbol{\beta}_0)^T - \boldsymbol{e}_j \left(\boldsymbol{e}_{2j-1}^T \left(1 - \pi(\boldsymbol{x}_j^T \boldsymbol{\beta}_0)\right) - \boldsymbol{e}_{2j}^T \pi(\boldsymbol{x}_j^T \boldsymbol{\beta}_0) \right) \right)$$

$$\times \tfrac{N}{N-n_j} \left(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\boldsymbol{\beta}_0) \right). \tag{14}$$

Now denoting $w_{j,N}(\boldsymbol{\beta}_0) \equiv \frac{n_j}{N} \pi(\boldsymbol{x}_j^T \boldsymbol{\beta}_0) \left(1 - \pi(\boldsymbol{x}_j^T \boldsymbol{\beta}_0)\right)$ the j-th diagonal element of $\boldsymbol{W}_N(\boldsymbol{\beta}_0)$ we can write

$$\left(\boldsymbol{X}_{(j)}^T \boldsymbol{W}_N^{(j)}(\boldsymbol{\beta}_0) \boldsymbol{X}_{(j)} \right)^{-1} = \tfrac{N-n_j}{N} \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0) \boldsymbol{X} - w_{j,N}(\boldsymbol{\beta}_0) \boldsymbol{x}_j \boldsymbol{x}_j^T \right)^{-1}$$

and

$$\tfrac{N}{N-n_j} \left(\boldsymbol{X}_{(j)}^T \boldsymbol{W}_N^{(j)}(\boldsymbol{\beta}_0) \boldsymbol{X}_{(j)} \right)^{-1} = \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0) \boldsymbol{X} \right)^{-1}$$

$$+ \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0) \boldsymbol{X} \right)^{-1} \tfrac{w_{j,N}(\boldsymbol{\beta}_0) \boldsymbol{x}_j \boldsymbol{x}_j^T}{1 - h_{jj,N}(\boldsymbol{\beta}_0)} \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0) \boldsymbol{X} \right)^{-1}, \tag{15}$$

where $h_{jj,N}(\boldsymbol{\beta}_0) \equiv w_{j,N}(\boldsymbol{\beta}_0)^{1/2} \boldsymbol{x}_j^T \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0) \boldsymbol{X} \right)^{-1} \boldsymbol{x}_j w_{j,N}(\boldsymbol{\beta}_0)^{1/2}$. Based on this last equality we have

$$\sqrt{N - n_j} (\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N} \left(\boldsymbol{0}_{(k+1)\times 1}, \boldsymbol{\Sigma}_{(j)}^{*} \right),$$

where

$$\boldsymbol{\Sigma}_{(j)}^{*} = (1 - \lambda_j) \left(\left(\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}_0) \boldsymbol{X} \right)^{-1} + \left(\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}_0) \boldsymbol{X} \right)^{-1} \tfrac{w_j(\boldsymbol{\beta}_0) \boldsymbol{x}_j \boldsymbol{x}_j^T}{1 - h_{jj}(\boldsymbol{\beta}_0)} \left(\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}_0) \boldsymbol{X} \right)^{-1} \right),$$

$w_j(\boldsymbol{\beta}_0) = \lim_{N\to\infty} w_{j,N}(\boldsymbol{\beta}_0)$ and $h_{jj}(\boldsymbol{\beta}_0) = \lim_{N\to\infty} h_{jj,N}(\boldsymbol{\beta}_0)$. Now we are going to express $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{(j)})$ as a linear combination of $\sqrt{N} \left(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\boldsymbol{\beta}_0) \right)$. Applying (14) and denoting

$$\boldsymbol{l}_j(\boldsymbol{\beta}_0) = (1 - \pi(\boldsymbol{x}_j^T \boldsymbol{\beta}_0)) \boldsymbol{e}_{2j-1} - \pi(\boldsymbol{x}_j^T \boldsymbol{\beta}_0) \boldsymbol{e}_{2j} \in \mathbb{R}^{2I}$$

we have

$$
\begin{aligned}
&\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0 \\
&= \left(\boldsymbol{X}_{(j)}^T \boldsymbol{W}_N^{(j)}(\boldsymbol{\beta}_0)\boldsymbol{X}_{(j)}\right)^{-1} \boldsymbol{X}^T \left(\bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)^T - \boldsymbol{e}_j \boldsymbol{l}_j(\boldsymbol{\beta}_0)^T\right) \tfrac{N}{N-n_j} \left(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\boldsymbol{\beta}_0)\right) + \mathrm{o}_P((N-n_j)^{-1/2}) \\
&= \tfrac{N-n_j}{N} \left(\left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} + \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \tfrac{w_{j,N}(\boldsymbol{\beta}_0)\boldsymbol{x}_j\boldsymbol{x}_j^T}{1-h_{jj,N}(\boldsymbol{\beta}_0)} \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \right) \\
&\times \boldsymbol{X}^T \left(\bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)^T - \boldsymbol{e}_j \boldsymbol{l}_j(\boldsymbol{\beta}_0)^T\right) \tfrac{N-n_j}{N} \left(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\boldsymbol{\beta}_0)\right) + \mathrm{o}_P\left(\tfrac{N-n_j}{N}(N-n_j)^{-1/2}\right)
\end{aligned}
$$

and after some algebra we can write

$$
\sqrt{N}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{(j)}) = \left(\boldsymbol{A}_N^{(j)}(\boldsymbol{\beta}_0) + \boldsymbol{B}_N^{(j)}(\boldsymbol{\beta}_0)\right) \sqrt{N}\left(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\boldsymbol{\beta}_0)\right) + \mathrm{o}_P(1),
$$

where

$$
\boldsymbol{A}_N^{(j)}(\boldsymbol{\beta}_0) = \tfrac{1}{1-h_{jj,N}(\boldsymbol{\beta}_0)} \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \boldsymbol{x}_j \boldsymbol{l}_j(\boldsymbol{\beta}_0)^T
$$

and

$$
\boldsymbol{B}_N^{(j)}(\boldsymbol{\beta}_0) = -\left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \tfrac{w_{j,N}(\boldsymbol{\beta}_0)\boldsymbol{x}_j\boldsymbol{x}_j^T}{1-h_{jj,N}(\boldsymbol{\beta}_0)} \left(\boldsymbol{X}^T \boldsymbol{W}_N(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)^T.
$$

In order to get the asymptotic distribution of $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{(j)})$ and taking into account (9) we need to obtain the following expressions:
i) $\overline{\boldsymbol{A}}^{(j)}(\boldsymbol{\beta}_0) \equiv \boldsymbol{A}^{(j)}(\boldsymbol{\beta}_0)\boldsymbol{\Sigma}\boldsymbol{A}^{(j)}(\boldsymbol{\beta}_0)^T$,
ii) $\overline{\boldsymbol{B}}^{(j)}(\boldsymbol{\beta}_0) \equiv \boldsymbol{B}^{(j)}(\boldsymbol{\beta}_0)\boldsymbol{\Sigma}\boldsymbol{B}^{(j)}(\boldsymbol{\beta}_0)^T$,
iii) $\overline{\boldsymbol{C}}^{(j)}(\boldsymbol{\beta}_0) \equiv \boldsymbol{B}^{(j)}(\boldsymbol{\beta}_0)\boldsymbol{\Sigma}\boldsymbol{A}^{(j)}(\boldsymbol{\beta}_0)^T$,
with $\boldsymbol{A}^{(j)}(\boldsymbol{\beta}_0) \equiv \lim_{N\to\infty} \boldsymbol{A}_N^{(j)}(\boldsymbol{\beta}_0)$ and $\boldsymbol{B}^{(j)}(\boldsymbol{\beta}_0) \equiv \lim_{N\to\infty} \boldsymbol{B}_N^{(j)}(\boldsymbol{\beta}_0)$, and $\boldsymbol{\Sigma}$ is given by (10).
In relation to i) we have

$$
\overline{\boldsymbol{A}}^{(j)}(\boldsymbol{\beta}_0) = \tfrac{1}{(1-h_{jj}(\boldsymbol{\beta}_0))^2} \left(\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \boldsymbol{x}_j \boldsymbol{l}_j(\boldsymbol{\beta}_0)^T \boldsymbol{\Sigma} \boldsymbol{l}_j(\boldsymbol{\beta}_0) \boldsymbol{x}_j^T \left(\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1},
$$

but $\boldsymbol{l}_j(\boldsymbol{\beta}_0)^T \boldsymbol{\Sigma} \boldsymbol{l}_j(\boldsymbol{\beta}_0) = \boldsymbol{c}_j^*(\boldsymbol{\beta}_0)^T \lambda_j \boldsymbol{\Sigma}_{\pi(\boldsymbol{x}_j^T\boldsymbol{\beta}_0)} \boldsymbol{c}_j^*(\boldsymbol{\beta}_0)$ and

$$
\begin{aligned}
\boldsymbol{c}_j^*(\boldsymbol{\beta}_0)^T \lambda_j \boldsymbol{\Sigma}_{\pi(\boldsymbol{x}_j^T\boldsymbol{\beta}_0)} \boldsymbol{c}_j^*(\boldsymbol{\beta}_0) &= \lambda_j \boldsymbol{c}_i^*(\boldsymbol{\beta}_0)^T \operatorname{diag}((\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0), 1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0))^T) \boldsymbol{c}_i^*(\boldsymbol{\beta}_0) \\
&\quad - \lambda_j \boldsymbol{c}_i^*(\boldsymbol{\beta}_0)^T (\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0), 1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0))^T (\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0), 1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)) \boldsymbol{c}_i^*(\boldsymbol{\beta}_0) \\
&= w_j(\boldsymbol{\beta}_0) \qquad (16)
\end{aligned}
$$

because

$$
\begin{aligned}
\boldsymbol{c}_i^*(\boldsymbol{\beta}_0)^T \operatorname{diag}((\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0), 1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0))^T) \boldsymbol{c}_i^*(\boldsymbol{\beta}_0) &= (\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0))(1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)), \\
\boldsymbol{c}_i^*(\boldsymbol{\beta}_0)^T (\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0), 1-\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0))^T &= 0.
\end{aligned}
$$

Therefore,

$$
\overline{\boldsymbol{A}}^{(j)}(\boldsymbol{\beta}_0) = \tfrac{1}{(1-h_{jj}(\boldsymbol{\beta}_0))^2} w_j(\boldsymbol{\beta}_0) \left(\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \boldsymbol{x}_j \boldsymbol{x}_j^T \left(\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}.
$$

In relation to ii) we have

$$\overline{\boldsymbol{B}}^{(j)}(\boldsymbol{\beta}_0) = \left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \tfrac{w_j(\boldsymbol{\beta}_0)^2\boldsymbol{x}_j\boldsymbol{x}_j^T}{(1-h_{jj}(\boldsymbol{\beta}_0))^2} \left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \left(\bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)^T\right)\boldsymbol{\Sigma} \\ \times \left(\bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)\right)\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^T\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}.$$

But taking into account (16)

$$\begin{aligned}\left(\bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)^T\right)\boldsymbol{\Sigma}\left(\bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)\right) &= \bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)^T\lambda_i\boldsymbol{\Sigma}_{\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)}c_i^*(\boldsymbol{\beta}_0)\\ &= \bigoplus_{i=1}^{I} w_i(\boldsymbol{\beta}_0) = \boldsymbol{W}(\boldsymbol{\beta}_0).\end{aligned}$$

Therefore,

$$\begin{aligned}\overline{\boldsymbol{B}}^{(j)}(\boldsymbol{\beta}_0) &= \tfrac{w_j(\boldsymbol{\beta}_0)}{(1-h_{jj}(\boldsymbol{\beta}_0))^2}\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{x}_j h_{jj}(\boldsymbol{\beta}_0)\boldsymbol{x}_j^T\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\\ &= w_j(\boldsymbol{\beta}_0)\tfrac{h_{jj}(\boldsymbol{\beta}_0)}{(1-h_{jj}(\boldsymbol{\beta}_0))^2}\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^T\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}.\end{aligned}$$

Finally, in relation to iii) we have

$$\overline{\boldsymbol{C}}^{(j)}(\boldsymbol{\beta}_0) = -\tfrac{1}{(1-h_{jj}(\boldsymbol{\beta}_0))^2}\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{x}_j\boldsymbol{l}_j(\boldsymbol{\beta}_0)^T\boldsymbol{\Sigma}\bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0) \\ \times\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^T w_j(\boldsymbol{\beta}_0)\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}.$$

But

$$\begin{aligned}\boldsymbol{l}_j(\boldsymbol{\beta}_0)^T\boldsymbol{\Sigma}\left(\bigoplus_{i=1}^{I} c_i^*(\boldsymbol{\beta}_0)\right)\boldsymbol{X} &= \boldsymbol{l}_j(\boldsymbol{\beta}_0)^T\left(\bigoplus_{i=1}^{I}\lambda_i\boldsymbol{\Sigma}_{\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)}c_i^*(\boldsymbol{\beta}_0)\boldsymbol{x}_i^T\right)\\ &= \boldsymbol{c}_j^*(\boldsymbol{\beta}_0)^T\lambda_j\boldsymbol{\Sigma}_{\pi(\boldsymbol{x}_j^T\boldsymbol{\beta}_0)}\boldsymbol{c}_j^*(\boldsymbol{\beta}_0)\boldsymbol{x}_j^T\\ &= w_j(\boldsymbol{\beta}_0)\boldsymbol{x}_j^T,\end{aligned}$$

and thus

$$\overline{\boldsymbol{C}}^{(j)}(\boldsymbol{\beta}_0) = -\tfrac{w_j(\boldsymbol{\beta}_0)h_{jj}(\boldsymbol{\beta}_0)}{(1-h_{jj}(\boldsymbol{\beta}_0))^2}\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^T\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}.$$

Then we have

$$\begin{aligned}&\left(\boldsymbol{A}^{(j)}(\boldsymbol{\beta}_0)+\boldsymbol{B}^{(j)}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{\Sigma}\left(\boldsymbol{A}^{(j)}(\boldsymbol{\beta}_0)+\boldsymbol{B}^{(j)}(\boldsymbol{\beta}_0)\right)\\ &= \tfrac{w_j(\boldsymbol{\beta}_0)}{(1-h_{jj}(\boldsymbol{\beta}_0))^2}\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^T\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\end{aligned}$$

and the result follows. ■

Once the asymptotic distribution of $\sqrt{N}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^{(j)})$ is known, the next step is to consider Cook's distance (5) as a quadratic form of $\sqrt{N}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^{(j)})$ defined by the inverse of the asymptotic variance-covariance matrix of $\sqrt{N}\widehat{\boldsymbol{\beta}}$. In the next theorem the asymptotic distribution of $D^{(j)}(\boldsymbol{\beta}_0)$ is established.

Theorem 2 *Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^{(j)}$ be the MLE's of parameters $\boldsymbol{\beta}$ based on the full observations and the full observations minus the j-th observation respectively. Then*

$$D^{(j)}(\boldsymbol{\beta}_0) \xrightarrow[N\to\infty]{\mathcal{L}} \frac{h_{jj}(\boldsymbol{\beta}_0)}{1-h_{jj}(\boldsymbol{\beta}_0)}\chi_1^2, \tag{17}$$

where

$$D^{(j)}(\boldsymbol{\beta}_0) \equiv N(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^{(j)})^T\mathbf{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\mathbf{X}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^{(j)}).$$

Proof. We shall use the following result "Suppose $(W_1,...,W_q)^T$, a $q\times 1$ random variable, is distributed as $\mathcal{N}(\mathbf{0}_{q\times 1},\mathbf{\Sigma}_W)$ and $\mathbf{M}$ is any real symmetric matrix of order q. Let $r=\text{rank}(\mathbf{\Sigma}_W\mathbf{M}\mathbf{\Sigma}_W)\geq 1$ and let λ_1,..., λ_r be non-zero eigenvalues of $\mathbf{M}\mathbf{\Sigma}_W$ $(r\leq q)$. Then $(W_1,...,W_q)\mathbf{M}(W_1,...,W_q)^T\sim\sum_{i=1}^r\lambda_i Z_i^2$ where $\{Z_i\}_{i=1}^r$are independent random variables so that $Z_i\sim\mathcal{N}(0,1)$, $i=1,...,r$" (see Dik and de Gunst [10, Corollary 2.1]).
Based on (13) we need to calculate

$$\text{rank}(\mathbf{\Sigma}^{(j)}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{(j)})=\text{rank}\left((\mathbf{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\mathbf{X})^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^T(\mathbf{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\mathbf{X})^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^T(\mathbf{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\mathbf{X})^{-1}\right).$$

Let $\mathbf{R}\equiv(\mathbf{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\mathbf{X})^{-1/2}\boldsymbol{x}_j$. Since $\text{rank}(\mathbf{R}\mathbf{R}^T\mathbf{R}\mathbf{R}^T)=\text{rank}(\mathbf{R}\mathbf{R}^T(\mathbf{R}\mathbf{R}^T)^T)=\text{rank}(\mathbf{R}\mathbf{R}^T)=\text{rank}(\mathbf{R})$ and $\mathbf{R}$ is a $(k+1)\times 1$ matrix, the rank of $\mathbf{\Sigma}^{(j)}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{(j)}$ is equal to $r=1$. On the other hand since the nonzero eigenvalues of $\mathbf{FG}$ are equal to the nonzero eigenvalues of $\mathbf{GF}$, when the dimensions of the $\mathbf{F}$ and $\mathbf{G}^T$ matrices are equal, we have that the nonzero eigenvalue of

$$\mathbf{\Sigma}^{(j)}\mathbf{\Sigma}^{-1}=\tfrac{w_j(\boldsymbol{\beta}_0)}{1-h_{jj}(\boldsymbol{\beta}_0)}\left(\mathbf{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\mathbf{X}\right)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^T=\mathbf{FG},$$

where $\mathbf{F}{=}w_j(\boldsymbol{\beta}_0)\left(\mathbf{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\mathbf{X}\right)^{-1}\boldsymbol{x}_j/(1{-}h_{jj}(\boldsymbol{\beta}_0))$ and $\mathbf{G}=\boldsymbol{x}_j^T$, coincides with the eigenvalue of

$$\mathbf{GF}=\frac{1}{1-h_{jj}(\boldsymbol{\beta}_0)}w_j(\boldsymbol{\beta}_0)\boldsymbol{x}_j^T\left(\mathbf{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\mathbf{X}\right)^{-1}\boldsymbol{x}_j=\frac{1}{1-h_{jj}(\boldsymbol{\beta}_0)}h_{jj}(\boldsymbol{\beta}_0).$$

Therefore, $h_{jj}(\boldsymbol{\beta}_0)/(1{-}h_{jj}(\boldsymbol{\beta}_0))$ is the nonzero eigenvalue of $\mathbf{\Sigma}^{(j)}\mathbf{\Sigma}^{-1}$ and we obtain the desired result. ■

To conclude this section, because according to Theorem 2 the statistic $D^{(j)}(\boldsymbol{\beta}_0)$ and also its asymptotic distribution depend on the same unknown parameter $\boldsymbol{\beta}_0$, we shall now establish the really useful result in the practice.

Corollary 3 *Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^{(j)}$ be the MLE's of parameters $\boldsymbol{\beta}$ based on the full observations and the full observations minus the j-th observation respectively. Then*

$$D^{(j)}(\widehat{\boldsymbol{\beta}}) \xrightarrow[N\to\infty]{\mathcal{L}} \frac{h_{jj}(\boldsymbol{\beta}_0)}{1-h_{jj}(\boldsymbol{\beta}_0)}\chi_1^2, \tag{18}$$

where $D^{(j)}(\widehat{\boldsymbol{\beta}})$ is given by (5).

Proof. It is straightforwardly obtained following Slutsky's Theorem in the same manner as in Ferguson [11, Lemma 2 of page 56] is establishes that Hottelling's T^2 statistic associated with an i.i.d. sample of d-dimensional random vectors is asymptotically Chi-squared with d degrees of freedom. ■

4 Looking for a distribution based cutoff for Cook's distance

Influence is the effective impact that an observation y_j has on the fit of the model. Traditionally, in logistic regression, one observation y_{j_1} is said to be more influential than another one y_{j_2}, $j_1, j_2 \in \{1, ..., I\}$, if $D^{(j_1)}(\widehat{\boldsymbol{\beta}}) > D^{(j_2)}(\widehat{\boldsymbol{\beta}})$. In order to classify an observation as influential, based on such idea jointly the result obtained through Theorem 2, arises the necessity of having a common cutoff once a significance level α is associated with the test of outliers through standardized residuals. However, the main inconvenience in (18) is that the quantile type cutoff of $D^{(j)}(\widehat{\boldsymbol{\beta}})$ depends on the index j, i.e. $h_{jj}(\boldsymbol{\beta}_0)\chi^2_{\alpha,1}/(1-h_{jj}(\boldsymbol{\beta}_0))$, where $\chi^2_{\alpha,1}$ is a quantile of order α for a χ^2_1 random variable, and thus it is not a valid cutoff for the rest of $D^{(i)}(\widehat{\boldsymbol{\beta}})$ values, $i = 1, ..., I$, $i \neq j$. In this section a cutoff for the Cook's distance is proposed which is in relationship with the criterions to classify an observations as outlier and leverage. Above all it should be remarked that once (18) is established, it is not possible to provide the usual quantile type cutoff given a significance level (for more details see Obenchain [19]), however it is possible to maintain as much as possible coherent criterion according to the asymptotic distribution of Cook's distance as well as the criterions chosen to classify outliers and leverage observations. We shall start revising some definitions in which we shall later base our criterion for a distribution based cutoff.

Definition 4 *The i-th Pearson residual, $r_i(\widehat{\boldsymbol{\beta}})$, $i = 1, ..., I$, is given by*

$$r_i(\widehat{\boldsymbol{\beta}}) = \sqrt{\frac{(y_i - n_i\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}))^2}{n_i\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})} + \frac{(n_i - y_i - n_i(1-\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})))^2}{n_i(1-\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}))}} = \sqrt{\frac{(y_i - n_i\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}))^2}{n_i\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})(1-\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}))}}. \tag{19}$$

In vector notation, $\boldsymbol{r}(\widehat{\boldsymbol{\beta}}) = (r_1(\widehat{\boldsymbol{\beta}}), ..., r_I(\widehat{\boldsymbol{\beta}}))^T$ is given as

$$\boldsymbol{r}(\widehat{\boldsymbol{\beta}}) = \left(\bigoplus_{i=1}^{I}\boldsymbol{J}_2^T\right)\operatorname{diag}\left(\boldsymbol{p}_N(\widehat{\boldsymbol{\beta}})\right)^{-\frac{1}{2}}\sqrt{N}(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\widehat{\boldsymbol{\beta}})) = \boldsymbol{W}(\widehat{\boldsymbol{\beta}})^{-\frac{1}{2}}\left(\bigoplus_{i=1}^{I}\boldsymbol{e}_1^T\right)\sqrt{N}(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\widehat{\boldsymbol{\beta}})).$$

where $\boldsymbol{J}_2^T = (1, 1)$ and $\boldsymbol{e}_1^T = (1, 0)$.

By knowing the asymptotic distribution of $\sqrt{N}(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\widehat{\boldsymbol{\beta}}))$, it is straightforwardly obtained

$$\boldsymbol{r}(\widehat{\boldsymbol{\beta}}) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}_{I\times 1}, \boldsymbol{I}_I - \boldsymbol{H}(\boldsymbol{\beta}_0)\right),$$

where $\boldsymbol{I}_I$ is $I \times I$ identity matrix and

$$\boldsymbol{H}(\boldsymbol{\beta}_0) \equiv \boldsymbol{W}(\boldsymbol{\beta}_0)^{\frac{1}{2}}\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{\beta}_0)^{\frac{1}{2}}$$

is a projection matrix called "Hat matrix". Since $\widehat{\boldsymbol{\beta}}$ is obtained as solution of (3), it holds

$$\boldsymbol{X}^T\left(\bigoplus_{i=1}^{I}\boldsymbol{e}_1^T\right)\sqrt{N}(\widehat{\boldsymbol{p}}_N - \boldsymbol{p}_N(\widehat{\boldsymbol{\beta}})) = \boldsymbol{0}_{I\times 1}$$

and hence

$$\boldsymbol{H}(\widehat{\boldsymbol{\beta}})\boldsymbol{r}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{0}_{I\times 1} \quad \text{and} \quad (\boldsymbol{I}_I - \boldsymbol{H}(\widehat{\boldsymbol{\beta}}))\boldsymbol{r}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{r}(\widehat{\boldsymbol{\beta}}).$$

This means that the projection matrix $\boldsymbol{I}_I - \boldsymbol{H}(\widehat{\boldsymbol{\beta}})$ spans the Pearson residuals space, as it happens in Linear Regression. By analogy with linear regression, the j-th diagonal element of the Hat matrix $\boldsymbol{H}(\widehat{\boldsymbol{\beta}})$, $h_{jj}(\widehat{\boldsymbol{\beta}})$ (defined in Theorem 1) is called "leverage". Even the work of Pregibon [21] is an extension of linear regression diagnostics to logistic regression, it is necessary to take care of the similarities among both models. Actually, while in linear regression fitted values are obtained multiplying the vector of observed values on the left side by the Hat matrix, i.e. h_{ii} is the amount of leverage per unit of y_i exerted on determining $\widehat{y}_i$, in logistic regression this is not so. This means that while in linear regression it holds $h_{ii} = \partial\widehat{y}_i/\partial y_i$, in logistics regression $h_{ii}(\boldsymbol{\beta}_0) \neq \partial(n_i\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}))/\partial y_i$. As it is mentioned in Hosmer and Lemeshow [14, page 171] some authors have questioned the appropriateness of $h_{ii}(\boldsymbol{\beta}_0)$ as leverage measure, for this reason we shall later propose a leverage measure based on a new concept of variability.

The next measure, defined by Pregibon [21], is an approximation of $D^{(j)}(\widehat{\boldsymbol{\beta}})$ and is also a usual influence measure in statistical packages.

Definition 5 *The Confidence Interval Displacement (CID) is given by*

$$\widetilde{D}^{(j)}(\widehat{\boldsymbol{\beta}}) \equiv \nu_j(\widehat{\boldsymbol{\beta}})r_j^*(\widehat{\boldsymbol{\beta}})^2, \tag{20}$$

where $r_j^*(\widehat{\boldsymbol{\beta}}) = r_j(\widehat{\boldsymbol{\beta}}) \Big/ \sqrt{1 - h_{jj}(\widehat{\boldsymbol{\beta}})}$ *is the standardized Pearson residual and*

$$\nu_j(\boldsymbol{\beta}_0) \equiv \frac{h_{jj}(\boldsymbol{\beta}_0)}{1 - h_{jj}(\boldsymbol{\beta}_0)}. \tag{21}$$

Observe that the asymptotic distribution of (20) is the same as (18). Let us interpret the term (21). By multiplying on the left side of (12) and (13) by $\boldsymbol{x}_j^T$, because $f(p)^{-1} = \text{logit}(p) \equiv \log(p/(1-p))$ is the inverse of $f(p) = e^p/(1-e^p)$, it is straightforwardly obtained

$$\begin{aligned}
&\sqrt{N}(\text{logit}(\pi(\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}})) - \text{logit}(\pi(\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}}^{(j)}))) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \tfrac{h_{jj}(\boldsymbol{\beta}_0)}{1-h_{jj}(\boldsymbol{\beta}_0)}\tfrac{h_{jj}(\boldsymbol{\beta}_0)}{w_{jj}(\boldsymbol{\beta}_0)}\right), \\
&\sqrt{N}(\text{logit}(\pi(\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}})) - \text{logit}(\pi(\boldsymbol{x}_j^T\boldsymbol{\beta}_0))) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \tfrac{h_{jj}(\boldsymbol{\beta}_0)}{w_{jj}(\boldsymbol{\beta}_0)}\right),
\end{aligned}$$

which means that the first term in the asymptotic distribution of (17), as well as in (20), represents the relative increment on the asymptotic variance of $\text{logit}(\pi(\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}}))$ when the j-th observation is omitted from the sample. By changing the interpretation of variability through a infinitesimal increment ($h_{ii} = \partial\widehat{y}_i/\partial y_i$), we shall now focus on the statistical interpretation of variability through the relative increment of the variance and therefore in the following we shall consider that (21) is a leverage measure. It should be noted that in this new framework a leverage observation remains having the potential to be influential but may not actually be so.

As traditionally, if $r_j^*(\widehat{\boldsymbol{\beta}}) > z_\alpha$, where z_α is a quantile of order α for a standard normal random variable, then the j-th observation is considered to be an outlier, i.e. anomalous observation for the fit of the model. Now we are going to give a new criterion for detecting leverage and influential observations.

Definition 6 *Once α significance level is prefixed for outliers, we propose the next criterion based on case-deletion diagnostics:*

- $2\bar{\nu}(\widehat{\boldsymbol{\beta}})$ *is a cutoff of* $\nu_j(\widehat{\boldsymbol{\beta}})$, *for considering the j-th observation to be a leverage, where*

$$\begin{aligned}
\bar{\nu}(\widehat{\boldsymbol{\beta}}) &= \frac{1}{I}\sum_{i=1}^{I}\nu_i(\widehat{\boldsymbol{\beta}}), \qquad \nu_i(\widehat{\boldsymbol{\beta}}) = \frac{h_{ii,N}(\widehat{\boldsymbol{\beta}})}{1 - h_{ii,N}(\widehat{\boldsymbol{\beta}})}, \\
h_{ii,N}(\widehat{\boldsymbol{\beta}}) &= w_{i,N}(\widehat{\boldsymbol{\beta}})^{1/2}\boldsymbol{x}_i^T\left(\boldsymbol{X}^T\boldsymbol{W}_N(\widehat{\boldsymbol{\beta}})\boldsymbol{X}\right)^{-1}\boldsymbol{x}_i w_{i,N}(\widehat{\boldsymbol{\beta}})^{1/2}, \\
w_{i,N}(\widehat{\boldsymbol{\beta}}) &= \frac{n_i}{N}\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})\left(1 - \pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})\right), \\
\boldsymbol{W}_N(\widehat{\boldsymbol{\beta}}) &= \text{diag}\left((w_{i,N}(\widehat{\boldsymbol{\beta}}))_{i=1,\ldots,I}\right);
\end{aligned}$$

• $\bar{\nu}(\widehat{\boldsymbol{\beta}})\chi^2_{\alpha,1}$ *is a cutoff of* $D^{(j)}(\widehat{\boldsymbol{\beta}})$ *(as well as* $\widetilde{D}^{(j)}(\widehat{\boldsymbol{\beta}})$*), for considering the* j*-th observation to be influential, where* $\chi^2_{\alpha,1}$ *is a quantile of order* α *for a* χ^2_1 *random variable.*

It must be stressed that

$$\nu_j(\boldsymbol{\beta}_0) > \bar{\nu}(\boldsymbol{\beta}_0) \Longrightarrow \Pr(D^{(j)}(\boldsymbol{\beta}_0) > \nu_j(\boldsymbol{\beta}_0)\chi^2_{\alpha,1}) = \alpha < \Pr(D^{(j)}(\boldsymbol{\beta}_0) > \bar{\nu}(\boldsymbol{\beta}_0)\chi^2_{\alpha,1}) = \Pr(Y_j \text{ influential}),$$
$$\nu_j(\boldsymbol{\beta}_0) < \bar{\nu}(\boldsymbol{\beta}_0) \Longrightarrow \Pr(D^{(j)}(\boldsymbol{\beta}_0) > \nu_j(\boldsymbol{\beta}_0)\chi^2_{\alpha,1}) = \alpha > \Pr(D^{(j)}(\boldsymbol{\beta}_0) > \bar{\nu}(\boldsymbol{\beta}_0)\chi^2_{\alpha,1}) = \Pr(Y_j \text{ influential}),$$

which means that a larger probability to be influential than α is assigned to those observations with larger leverage measure than the average. Thus, the first property of the criterion above is the higher sensitivity for those observations which have larger leverage measure (they are potentially more influential). On the other hand, if we increase the sensitivity for detecting outliers by reducing α value, we also increase the sensitivity for detecting influential observations.

It should be also pointed out the reason why the constant of proportionality of $\bar{\nu}(\widehat{\boldsymbol{\beta}})$ is chosen to be 2 for the cutoffs associated with leverage observations and 1 for the cutoffs associated with influential observations. Such a calibration for a leverage and also for an influential observation, permits us to establish the following properties

$$y_j \text{ leverage and } y_j \text{ outlier} \Rightarrow y_j \text{ influential}, \tag{22}$$
$$y_j \text{ influential} \not\Rightarrow y_j \text{ leverage or } y_j \text{ outlier}, \tag{23}$$

actually when the j-th observation is non leverage being $\bar{\nu}(\widehat{\boldsymbol{\beta}}) \leq \nu_j(\widehat{\boldsymbol{\beta}}) \leq 2\bar{\nu}(\widehat{\boldsymbol{\beta}})$, such an observation could be influential and not outlier. Even though (23) is accepted in a theoretical framework, it has never been considered any calibration based on such a property for none influence measure as we know, and this is mainly due to the lack of distribution based cutoff values associated with influence measures.

Since according to the criterion given in Definition 6 we have

$$p(j,\alpha) \equiv \Pr(Y_j \text{ influential}) = \Pr(\chi^2_1 > \tfrac{\bar{\nu}(\boldsymbol{\beta}_0)}{\nu_j(\boldsymbol{\beta}_0)}\chi^2_{\alpha,1}), \tag{24}$$

and taking into account that $p(j,\alpha)$ is unknown, because $\boldsymbol{\beta}_0$ is unknown, we can define the approximated probability of (24)

$$\widehat{p}(j,\alpha) \equiv \Pr(Y_j \text{ influential}) = \Pr(\chi^2_1 > \tfrac{\bar{\nu}(\widehat{\boldsymbol{\beta}})}{\nu_j(\widehat{\boldsymbol{\beta}})}\chi^2_{\alpha,1}).$$

Finally, it should be stressed that although a leverage observation is not probabilistically measurable because $x_{i0},...,x_{ik}$ are not random coefficients ($i \in \{1,...,I\}$), according to the criterion given in Definition 6 and taking into account Definition 5 we have

$$\Pr(Y_j \text{ influential}|Y_j \text{ outlier}) \simeq \Pr\Big(\tfrac{D^{(j)}(\widehat{\boldsymbol{\beta}})}{\nu_j(\widehat{\boldsymbol{\beta}})} > \tfrac{\bar{\nu}(\widehat{\boldsymbol{\beta}})}{\nu_j(\widehat{\boldsymbol{\beta}})}\chi^2_{\alpha,1} \Big| \tfrac{D^{(j)}(\widehat{\boldsymbol{\beta}})}{\nu_j(\widehat{\boldsymbol{\beta}})} > \chi^2_{\alpha,1}\Big) \simeq \begin{cases} \frac{\widehat{p}(j,\alpha)}{\alpha}, & \nu_j(\widehat{\boldsymbol{\beta}}) < \bar{\nu}(\widehat{\boldsymbol{\beta}}) \\ 1, & \nu_j(\widehat{\boldsymbol{\beta}}) \geq \bar{\nu}(\widehat{\boldsymbol{\beta}}) \end{cases},$$

$$\Pr(Y_j \text{ outlier}|Y_j \text{ influential}) \simeq \Pr\Big(\tfrac{D^{(j)}(\widehat{\boldsymbol{\beta}})}{\nu_j(\widehat{\boldsymbol{\beta}})} > \chi^2_{\alpha,1} \Big| \tfrac{D^{(j)}(\widehat{\boldsymbol{\beta}})}{\nu_j(\widehat{\boldsymbol{\beta}})} > \tfrac{\bar{\nu}(\widehat{\boldsymbol{\beta}})}{\nu_j(\widehat{\boldsymbol{\beta}})}\chi^2_{\alpha,1}\Big) \simeq \begin{cases} 1, & \nu_j(\widehat{\boldsymbol{\beta}}) < \bar{\nu}(\widehat{\boldsymbol{\beta}}) \\ \frac{\alpha}{\widehat{p}(j,\alpha)}, & \nu_j(\widehat{\boldsymbol{\beta}}) \geq \bar{\nu}(\widehat{\boldsymbol{\beta}}) \end{cases},$$

and

$$\Pr(Y_j \text{ influential and } Y_j \text{ outlier}) \simeq \begin{cases} \widehat{p}(j,\alpha) < \alpha, & \nu_j(\widehat{\boldsymbol{\beta}}) < \bar{\nu}(\widehat{\boldsymbol{\beta}}) \\ \alpha, & \nu_j(\widehat{\boldsymbol{\beta}}) \geq \bar{\nu}(\widehat{\boldsymbol{\beta}}) \end{cases}.$$

5 Numerical Examples

Case-deletion diagnostics for logistic regression models are illustrated by examining four data sets. The method used for determining influence on the model assumes that the chosen model is suitable and any observation that have large influence will affect the analysis of the model. A usual tool for detecting influential observations is the index plot (sometimes called dot plot), which consists in plotting a diagnostic measure against the case order, to visually inspect the influence of a case-deletion on the individuals. We shall graphic such plots simultaneously for Cook's distance, the leverage measure (21) and standardized residuals and at the same time we shall analyze whether the conclusions obtained visually through index plots would differ in comparison with the cutoffs proposed in the criterion given in Definition 6. We also shall focus in comparing the numerical accuracy of the approximation of Cook's distance $D^{(j)}(\widehat{\boldsymbol{\beta}})$, through $\widetilde{D}^{(j)}(\widehat{\boldsymbol{\beta}})$ given in Definition 5.

Example 1: Finney (1947). Pregibon [21] and Finney [12] studied the data shown in Table 1 where the interest is focussed on the occurrence ($y_i = 1$) or nonoccurrence ($y_i = 0$) of vasoconstriction in the skin of the finger as a function of the logarithm of volume (x'_{i1}) and rate (x'_{i2}) of inspired air measured in litres. The model, the estimated values of the parameter as well as the goodness-of-fit test statistic are given as follows

$$\text{logit}(\pi(\boldsymbol{x}_i^T\widehat{\beta})) = \widehat{\beta}_0 + x_{i1}\widehat{\beta}_1 + x_{i2}\widehat{\beta}_2, \qquad x_{ij} = \log x'_{ij}, \qquad i = 1,...,I = 39, \quad j = 1,2,$$
$$\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)^T = (-2.8754, 5.1793, 4.5617)^T, \qquad k = 2,$$
$$X^2(\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{I} r_i^*(\widehat{\boldsymbol{\beta}})^2 = 34.2338, \qquad \text{p-value}(X^2(\widehat{\boldsymbol{\beta}})) = P(\chi^2_{I-k-1} > X^2(\widehat{\boldsymbol{\beta}})) = 0.5528.$$

And the estimated probabilities and diagnostics associated with the $I = 39$ individuals are given in Table 2. The corresponding index plot is in part above of Figure 1 where taking into account the criterion given in Definition 6 and the thresholds shown in Table 9, remarkable observations are highlighted through diamond symbols.

i	x'_{i1}	x'_{i2}	y_i	n_i	i	x'_{i1}	x'_{i2}	y_i	n_i	i	x'_{i1}	x'_{i2}	y_i	n_i
1	3.70	0.825	1	1	14	1.40	2.330	1	1	27	1.80	1.500	1	1
2	3.50	1.090	1	1	15	0.75	3.750	1	1	28	0.95	1.900	0	1
3	1.25	2.500	1	1	16	2.30	1.640	1	1	29	1.90	0.950	1	1
4	0.75	1.500	1	1	17	3.20	1.600	1	1	30	1.60	0.400	0	1
5	0.80	3.200	1	1	18	0.85	1.415	1	1	31	2.70	0.750	1	1
6	0.70	3.500	1	1	19	1.70	1.060	0	1	32	2.35	0.030	0	1
7	0.60	0.750	0	1	20	1.80	1.800	1	1	33	1.10	1.830	0	1
8	1.10	1.700	0	1	21	0.40	2.000	0	1	34	1.10	2.200	1	1
9	0.90	0.750	0	1	22	0.95	1.360	0	1	35	1.20	2.000	1	1
10	0.90	0.450	0	1	23	1.35	1.350	0	1	36	0.80	3.330	1	1
11	0.80	0.570	0	1	24	1.50	1.360	0	1	37	0.95	1.900	0	1
12	0.55	2.750	0	1	25	1.60	1.780	1	1	38	0.75	1.900	0	1
13	0.60	3.000	0	1	26	0.60	1.500	0	1	39	1.30	1.625	1	1

Table 1: Data from Finney (1947).

i	$\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})$	$r_i^*(\widehat{\boldsymbol{\beta}})$	$h_{ii}(\widehat{\boldsymbol{\beta}})$	$\nu_i(\widehat{\boldsymbol{\beta}})$	$\widehat{p}(i, 0.05)$	$\widetilde{D}^{(i)}(\widehat{\boldsymbol{\beta}})$	$D^{(i)}(\widehat{\boldsymbol{\beta}})$
1	0.9536	0.2315	0.0927	0.1021	0.0706	0.0055	0.0053
2	0.9821	0.1379	0.0429	0.0448	0.0063	0.0009	0.0008
3	0.9213	0.3016	0.0612	0.0652	0.0237	0.0059	0.0057
4	0.0748	3.6814	0.0867	0.0950	0.0609	1.2873	3.5603
5	0.7816	0.5622	0.1158	0.1310	0.1104	0.0414	0.0386
6	0.7295	0.6615	0.1524	0.1798	0.1730	0.0787	0.0734
7	0.0011	−0.0329	0.0076	0.0077	0.0000	0.0000	0.0000
8	0.5097	−1.0493	0.0559	0.0592	0.0176	0.0652	0.0640
9	0.0087	−0.0954	0.0342	0.0354	0.0021	0.0003	0.0003
10	0.0009	−0.0294	0.0072	0.0073	0.0000	0.0000	0.0000
11	0.0014	−0.0371	0.0097	0.0098	0.0000	0.0000	0.0000
12	0.2047	−0.5496	0.1481	0.1738	0.1658	0.0525	0.0486
13	0.3753	−0.8470	0.1628	0.1944	0.1901	0.1395	0.1346
14	0.9385	0.2633	0.0551	0.0583	0.0167	0.0040	0.0039
15	0.8408	0.4675	0.1336	0.1542	0.1412	0.0337	0.0316
16	0.9758	0.1609	0.0402	0.0419	0.0048	0.0011	0.0011
17	0.9950	0.0715	0.0172	0.0175	0.0000	0.0001	0.0001
18	0.1059	3.0555	0.0954	0.1054	0.0751	0.9845	2.2026
19	0.5346	−1.1501	0.1315	0.1514	0.1376	0.2003	0.2143
20	0.9453	0.2471	0.0525	0.0554	0.0141	0.0034	0.0033
21	0.0114	−0.1096	0.0373	0.0387	0.0033	0.0005	0.0005
22	0.1495	−0.4423	0.1015	0.1129	0.0855	0.0221	0.0202
23	0.5120	−1.0656	0.0761	0.0824	0.0442	0.0935	0.0943
24	0.6519	−1.4203	0.0717	0.0773	0.0377	0.1558	0.1745
25	0.8993	0.3449	0.0587	0.0623	0.0206	0.0074	0.0071
26	0.0248	−0.1640	0.0548	0.0579	0.0164	0.0016	0.0015
27	0.8827	0.3772	0.0661	0.0708	0.0299	0.0101	0.0096
28	0.4469	−0.9295	0.0647	0.0692	0.0281	0.0597	0.0570
29	0.5535	0.9847	0.1682	0.2022	0.1989	0.1961	0.1874
30	0.0097	−0.1018	0.0507	0.0534	0.0124	0.0006	0.0005
31	0.7224	0.7138	0.2459	0.3261	0.3117	0.1661	0.1549
32	5×10^{-7}	−0.0007	0.0000	0.0000	0.0000	0.0000	0.0000
33	0.5926	−1.2382	0.0510	0.0538	0.0127	0.0824	0.0848
34	0.7712	0.5619	0.0601	0.0640	0.0224	0.0202	0.0191
35	0.7740	0.5560	0.0552	0.0584	0.0168	0.0180	0.0171
36	0.8110	0.5140	0.1177	0.1333	0.1136	0.0352	0.0329
37	0.4469	−0.9295	0.0647	0.0692	0.0281	0.0597	0.0570
38	0.1919	−0.5137	0.1000	0.1112	0.0832	0.0293	0.0267
39	0.6678	0.7248	0.0531	0.0561	0.0147	0.0295	0.0284

Table 2: Diagnostics in data from Finney (1947).

Example 2: Brown (1980). Zelterman [24, Section 3.3] and Brown [6] studied the data shown in Table 3 where the interest is focussed on the evidence of lymphatic cancer ($y_i = 1$) or non evidence of lymphatic cancer ($y_i = 0$) in prostate cancer patients for predicting lymph nodal involvement of cancer as a function of five covariates (three dichotomous and two continuous): the X-ray finding ($x_{i1} = 1$ (presence), $x_{i1} = 0$ (absence)), size of the tumor by palpation ($x_{i2} = 1$ (serious), $x_{i2} = 0$ (non serious)), pathology grade by biopsy ($x_{i3} = 1$ (serious), $x_{i3} = 0$ (non serious)), the age of the patient at the time of diagnosis (x_{i4}) and serum acid phosphatase level (x_{i5}). The model, the estimated values of the parameter as well as the goodness-of-fit test statistic are given as follows

$$\mathrm{logit}(\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})) = \widehat{\beta}_0 + x_{i1}\widehat{\beta}_1 + x_{i2}\widehat{\beta}_2 + x_{i3}\widehat{\beta}_3 + x_{i4}\widehat{\beta}_4 + x_{i5}\widehat{\beta}_5, \qquad i = 1, ..., I = 53,$$
$$\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4, \widehat{\beta}_5)^T = (0.0618, 2.0453, 1.5641, 0.7614, 0.0692, 0.0243)^T, \qquad k = 5,$$
$$X^2(\widehat{\boldsymbol{\beta}}) = \textstyle\sum_{i=1}^{I} r_i^*(\widehat{\boldsymbol{\beta}})^2 = 46.7905, \qquad \text{p-value}(X^2(\widehat{\boldsymbol{\beta}})) = P(\chi^2_{I-k-1} > X^2(\widehat{\boldsymbol{\beta}})) = 0.4812,$$

and the estimated probabilities and diagnostics associated with the $I = 53$ individuals are given in Table 4. The corresponding index plot is in part below of Figure 1 where taking into account the criterion given in Definition 6 and the thresholds shown in Table 9, remarkable observations are highlighted through diamond symbols.

i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	y_i	n_i	i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	y_i	n_i
1	0	0	0	66	48	0	1	28	0	1	0	61	50	0	1
2	0	0	0	68	56	0	1	29	0	1	1	64	50	0	1
3	0	0	0	66	50	0	1	30	0	1	0	63	40	0	1
4	0	0	0	56	52	0	1	31	0	1	1	52	55	0	1
5	0	0	0	58	50	0	1	32	0	1	1	66	59	0	1
6	0	0	0	60	49	0	1	33	1	1	0	58	48	1	1
7	1	0	0	65	46	0	1	34	1	1	1	57	51	1	1
8	1	0	0	60	62	0	1	35	0	1	0	65	49	1	1
9	0	0	1	50	56	1	1	36	0	1	1	65	48	0	1
10	1	0	0	49	55	0	1	37	1	1	1	59	63	0	1
11	0	0	0	61	62	0	1	38	0	1	0	61	102	0	1
12	0	0	0	58	71	0	1	39	0	1	0	53	76	0	1
13	0	0	0	51	65	0	1	40	0	1	0	67	95	0	1
14	1	0	1	67	67	1	1	41	0	1	1	53	66	0	1
15	0	0	1	67	47	0	1	42	1	1	1	65	84	1	1
16	0	0	0	51	49	0	1	43	1	1	1	50	81	1	1
17	0	0	1	56	50	0	1	44	1	1	1	60	76	1	1
18	0	0	0	60	78	0	1	45	0	1	1	45	70	1	1
19	0	0	0	52	83	0	1	46	1	1	1	56	78	1	1
20	0	0	0	56	98	0	1	47	0	1	0	46	70	1	1
21	0	0	0	67	52	0	1	48	0	1	0	67	67	1	1
22	0	0	0	63	75	0	1	49	0	1	0	63	82	1	1
23	0	0	1	59	99	1	1	50	0	1	1	57	67	1	1
24	0	0	0	64	187	0	1	51	1	1	0	51	72	1	1
25	1	0	0	61	136	1	1	52	1	1	0	64	89	1	1
26	0	0	0	56	82	1	1	53	1	1	1	68	126	1	1
27	0	1	1	64	40	0	1								

Table 3: Data from Brown (1980).

i	$\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})$	$r_i^*(\widehat{\boldsymbol{\beta}})$	$h_{ii}(\widehat{\boldsymbol{\beta}})$	$\nu_i(\widehat{\boldsymbol{\beta}})$	$\widehat{p}(i, 0.05)$	$\widetilde{D}^{(i)}(\widehat{\boldsymbol{\beta}})$	$D^{(i)}(\widehat{\boldsymbol{\beta}})$
1	0.0342	–0.1911	0.0302	0.0311	0.0000	0.0011	0.0011
2	0.0361	–0.1968	0.0327	0.0338	0.0001	0.0013	0.0013
3	0.0358	–0.1958	0.0306	0.0316	0.0000	0.0012	0.0012
4	0.0724	–0.2852	0.0409	0.0426	0.0003	0.0035	0.0034
5	0.0608	–0.2590	0.0359	0.0372	0.0001	0.0025	0.0024
6	0.0521	–0.2384	0.0328	0.0339	0.0001	0.0019	0.0019
7	0.2184	–0.5805	0.1707	0.2058	0.1011	0.0694	0.0676
8	0.3684	–0.8352	0.1638	0.1959	0.0928	0.1366	0.1340
9	0.2182	2.0680	0.1620	0.1933	0.0907	0.8265	0.9728
10	0.5131	–1.1961	0.2634	0.3576	0.2135	0.5116	0.5183
11	0.0658	–0.2699	0.0339	0.0351	0.0001	0.0026	0.0025
12	0.0974	–0.3353	0.0405	0.0422	0.0003	0.0047	0.0046
13	0.1315	–0.4042	0.0733	0.0791	0.0082	0.0129	0.0125
14	0.4649	1.2616	0.2767	0.3826	0.2291	0.6089	0.6344
15	0.0646	–0.2722	0.0685	0.0735	0.0061	0.0054	0.0054
16	0.0930	–0.3311	0.0645	0.0690	0.0046	0.0076	0.0074
17	0.1373	–0.4193	0.0949	0.1048	0.0216	0.0184	0.0179
18	0.1002	–0.3407	0.0406	0.0424	0.0003	0.0049	0.0048
19	0.1796	–0.4886	0.0831	0.0907	0.0135	0.0217	0.0209
20	0.1929	–0.5089	0.0767	0.0831	0.0099	0.0215	0.0210
21	0.0351	–0.1939	0.0313	0.0323	0.0000	0.0012	0.0012
22	0.0776	–0.2957	0.0379	0.0394	0.0002	0.0034	0.0034
23	0.2988	1.6876	0.1760	0.2136	0.1075	0.6082	0.6841
24	0.5453	–1.6520	0.5606	1.2758	0.5102	3.4817	4.3771
25	0.7673	0.6177	0.2053	0.2583	0.1433	0.0986	0.0958
26	0.1394	2.5549	0.0540	0.0571	0.0019	0.3725	0.4664
27	0.2551	–0.6215	0.1135	0.1280	0.0376	0.0494	0.0475
28	0.2007	–0.5239	0.0849	0.0928	0.0146	0.0255	0.0247
29	0.3041	–0.7012	0.1114	0.1254	0.0357	0.0617	0.0592
30	0.1463	–0.4332	0.0865	0.0947	0.0156	0.0178	0.0173
31	0.5311	–1.1427	0.1324	0.1526	0.0569	0.1993	0.2085
32	0.3214	–0.7363	0.1265	0.1449	0.0507	0.0785	0.0754
33	0.6948	0.7313	0.1788	0.2177	0.1109	0.1164	0.1120
34	0.8490	0.4436	0.0960	0.1062	0.0225	0.0209	0.0201
35	0.1567	2.4307	0.0891	0.0978	0.0174	0.5776	0.7175
36	0.2797	–0.6631	0.1170	0.1325	0.0410	0.0583	0.0559
37	0.8676	–2.6734	0.0830	0.0906	0.0135	0.6473	0.9003
38	0.4711	–1.0146	0.1349	0.1559	0.0596	0.1605	0.1591
39	0.4515	–0.9714	0.1277	0.1464	0.0519	0.1381	0.1402
40	0.3314	–0.7629	0.1482	0.1740	0.0745	0.1013	0.0988
41	0.5801	–1.2544	0.1219	0.1388	0.0459	0.2183	0.2311
42	0.8782	0.3915	0.0955	0.1056	0.0221	0.0162	0.0155
43	0.9499	0.2363	0.0551	0.0583	0.0021	0.0033	0.0032
44	0.8935	0.3585	0.0729	0.0786	0.0080	0.0101	0.0097
45	0.7261	0.6897	0.2069	0.2608	0.1452	0.1241	0.1215
46	0.9208	0.3027	0.0611	0.0651	0.0036	0.0060	0.0058
47	0.5360	1.0626	0.2332	0.3042	0.1774	0.3435	0.3360
48	0.2005	2.1101	0.1042	0.1164	0.0292	0.5182	0.6248

49	0.3227	1.5285	0.1019	0.1134	0.0272	0.2650	0.2912
50	0.5176	1.0179	0.1007	0.1119	0.0262	0.1160	0.1160
51	0.8690	0.4128	0.1149	0.1298	0.0390	0.0221	0.0215
52	0.8030	0.5321	0.1337	0.1543	0.0583	0.0437	0.0420
53	0.9422	0.2594	0.0874	0.0958	0.0162	0.0064	0.0063

Table 4: Diagnostics in data from Brown (1980).

Example 3: Feigl and Zelen (1965). Muñoz-García et al. [18], Jonhson [16], Cook and Weisberg [8, page 185] and Feigl and Zelen [13] studied the data, shown in Table 5, on survival ($y_i = 1$) or non survival ($y_i = 0$) of $N = 33$ leukemia patients along 52 weeks as a function of a certain morphological characteristic in the white cells finding ($x_{i1} = 1$ (presence), $x_{i1} = 0$ (absence)) and their white blood cell count (x_{i2}). The model, the estimated values of the parameter as well as the goodness-of-fit test statistic are given as follows

$$\begin{aligned}
&\text{logit}(\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})) = \widehat{\beta}_0 + x_{i1}\widehat{\beta}_1 + x_{i2}\widehat{\beta}_2, \qquad i = 1, ..., I = 30,\\
&\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)^T = (-1.3073, 2.2610, -0.0317)^T, \qquad k = 2,\\
&X^2(\widehat{\boldsymbol{\beta}}) = \textstyle\sum_{i=1}^{I} r_i^*(\widehat{\boldsymbol{\beta}})^2 = 23.9798, \qquad \text{p-value}(X^2(\widehat{\boldsymbol{\beta}})) = P(\chi^2_{I-k-1} > X^2(\widehat{\boldsymbol{\beta}})) = 0.6314.
\end{aligned}$$

The estimated probabilities and diagnostics associated with the $I = 30$ observations are given in Table 6. The corresponding index plot is in part above of Figure 2 where taking into account the criterion given in Definition 6 and the thresholds shown in Table 9, remarkable observations are highlighted through diamond symbols.

i	x_{i1}	x_{i2}	y_i	n_i	i	x_{i1}	x_{i2}	y_i	n_i	i	x_{i1}	x_{i2}	y_i	n_i
1	1	2.30	1	1	11	1	9.40	1	1	21	0	5.30	0	1
2	1	0.75	1	1	12	1	32.00	0	1	22	0	10.00	0	1
3	1	4.30	1	1	13	1	35.00	0	1	23	0	19.00	0	1
4	1	2.60	1	1	14	1	52.00	0	1	24	0	27.00	0	1
5	1	6.00	0	1	15	1	100.00	1	3	25	0	28.00	0	1
6	1	10.50	1	1	16	0	4.40	1	1	26	0	31.00	0	1
7	1	10.00	1	1	17	0	3.00	1	1	27	0	26.00	0	1
8	1	17.00	0	1	18	0	4.00	0	1	28	0	21.00	0	1
9	1	5.40	0	1	19	0	1.50	0	1	29	0	79.00	0	1
10	1	7.00	1	1	20	0	9.00	0	1	30	0	100.00	0	2

Table 5: Data from Feigl and Zelen (1965).

i	$\pi(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})$	$r_i^*(\widehat{\boldsymbol{\beta}})$	$h_{ii}(\widehat{\boldsymbol{\beta}})$	$\nu_i(\widehat{\boldsymbol{\beta}})$	$\widehat{p}(i, 0.05)$	$\widetilde{D}^{(i)}(\widehat{\boldsymbol{\beta}})$	$D^{(i)}(\widehat{\boldsymbol{\beta}})$
1	0.7070	0.6737	0.0868	0.0950	0.0147	0.0431	0.0414
2	0.7171	0.6582	0.0892	0.0980	0.0163	0.0425	0.0407
3	0.6936	0.6943	0.0837	0.0914	0.0129	0.0440	0.0423
4	0.7050	0.6768	0.0863	0.0944	0.0144	0.0432	0.0415
5	0.6820	–1.5280	0.0813	0.0885	0.0115	0.2067	0.2299
6	0.6502	0.7630	0.0761	0.0824	0.0088	0.0480	0.0462
7	0.6538	0.7572	0.0766	0.0830	0.0090	0.0476	0.0458
8	0.6020	–1.2767	0.0722	0.0778	0.0070	0.1269	0.1347
9	0.6861	–1.5433	0.0822	0.0895	0.0119	0.2132	0.2378
10	0.6751	0.7233	0.0800	0.0870	0.0108	0.0455	0.0437
11	0.6582	0.7503	0.0772	0.0837	0.0093	0.0471	0.0453
12	0.4843	–1.0125	0.0840	0.0917	0.0130	0.0940	0.0891
13	0.4605	–0.9685	0.0899	0.0988	0.0167	0.0927	0.0855
14	0.3322	–0.7600	0.1388	0.1612	0.0610	0.0931	0.0756
15	0.0977	2.3200	0.6487	1.8467	0.5800	9.9398	143.6078
16	0.1904	2.1693	0.0967	0.1070	0.0215	0.5035	0.7902
17	0.1974	2.1266	0.1009	0.1123	0.0248	0.5077	0.7889
18	0.1924	–0.5139	0.0979	0.1085	0.0224	0.0286	0.0266
19	0.2051	–0.5371	0.1059	0.1184	0.0288	0.0342	0.0316
20	0.1689	–0.4712	0.0847	0.0925	0.0134	0.0205	0.0192
21	0.1861	–0.5023	0.0941	0.1038	0.0196	0.0262	0.0243
22	0.1645	–0.4632	0.0825	0.0899	0.0121	0.0193	0.0180
23	0.1289	–0.3983	0.0676	0.0725	0.0052	0.0115	0.0109
24	0.1029	–0.3494	0.0602	0.0640	0.0030	0.0078	0.0075
25	0.1000	–0.3438	0.0595	0.0633	0.0028	0.0075	0.0072
26	0.0918	–0.3275	0.0578	0.0613	0.0024	0.0066	0.0063
27	0.1059	–0.3551	0.0609	0.0648	0.0031	0.0082	0.0078
28	0.1219	–0.3854	0.0654	0.0699	0.0044	0.0104	0.0099
29	0.0215	–0.1516	0.0433	0.0452	0.0004	0.0010	0.0010
30	0.0112	–0.1557	0.0696	0.0748	0.0060	0.0018	0.0018

Table 6: Diagnostics in data from Feigl and Zelen (1965).

Example 4: Bickel et al. (1975). Table 7, taken from Bickel et al. [5], came from a study of the effect of applicant's gender on whether admitted into graduate school at Berkeley University in 1973. This example is the traditional one for studying diagnostics in loglinear models (see for instance Agresti [1, Section 7.3 in the first Edition (1990)]). Loglinear model is related to logistic model, as is explained in Agresti [2, Section 6.5] in such a way that the admission is a response variable, and covariates, the sex (S) and the department (D), for the logistic model should be understand through the proposed loglinear model. We shall focus on the loglinear model $\log(\pi(\widetilde{\boldsymbol{x}}_{ijk}^T\widehat{\boldsymbol{\theta}})) = \widehat{\theta}_0 + \widehat{\theta}_i^S + \widehat{\theta}_j^D + \widehat{\theta}_k^A + \widehat{\theta}_{ij}^{SD} + \widehat{\theta}_{ik}^{SA} + \widehat{\theta}_{jk}^{DA}$ whose corresponding logistic model is given by

$$
\begin{aligned}
&\operatorname{logit}(\pi(\boldsymbol{x}_{ij}^T\widehat{\boldsymbol{\beta}})) = \widehat{\beta}_0 + \widehat{\beta}_i^S + \widehat{\beta}_j^D, \quad i = 1, 2,\ j = 1, ..., 6, \quad I = 2 \times 6 = 12,\\
&\textstyle\sum_{i=1}^{2}\widehat{\beta}_i^S = \sum_{j=1}^{6}\widehat{\beta}_j^D = 0 \quad \text{(restrictions to avoid overparametrization)},\\
&\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1^S, \widehat{\beta}_1^D, \widehat{\beta}_2^D, \widehat{\beta}_3^D, \widehat{\beta}_4^D, \widehat{\beta}_5^D)^T\\
&\quad = (-0.6424, -0.0499, 1.2744, 1.2310, 0.0118, -0.0202, -0.4649)^T, \qquad k = 6,\\
&X^2(\widehat{\boldsymbol{\beta}}) = \textstyle\sum_{i=1}^{2}\sum_{j=1}^{6} r_{ij}^*(\widehat{\boldsymbol{\beta}})^2 = 18.8243, \qquad \text{p-value}(X^2(\widehat{\boldsymbol{\beta}})) = P(\chi^2_{I-k-1} > X^2(\widehat{\boldsymbol{\beta}})) = 0.0020.
\end{aligned}
$$

	Admitted			
	Male		Female	
Department	Yes	No	Yes	No
A	512	313	89	19
B	353	207	17	8
C	120	205	202	391
D	138	279	131	244
E	53	138	94	299
F	22	351	24	317

Table 7: Data from Bickel et al. (1975).

Therefore, the design matrix of the model is given by

$$
\boldsymbol{X} = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0\\
1 & 1 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 0 & 0 & 1 & 0 & 0\\
1 & 1 & 0 & 0 & 0 & 1 & 0\\
1 & 1 & 0 & 0 & 0 & 0 & 1\\
1 & 1 & -1 & -1 & -1 & -1 & -1\\
1 & -1 & 1 & 0 & 0 & 0 & 0\\
1 & -1 & 0 & 1 & 0 & 0 & 0\\
1 & -1 & 0 & 0 & 1 & 0 & 0\\
1 & -1 & 0 & 0 & 0 & 1 & 0\\
1 & -1 & 0 & 0 & 0 & 0 & 1\\
1 & -1 & -1 & -1 & -1 & -1 & -1
\end{pmatrix},
$$

which is in correspondence with the lexicographical order chosen for the indices of the vector of observed frequencies as well as the vector of trials

$$\begin{aligned}
\boldsymbol{y} &= (y_{11}, y_{21}, y_{31}, y_{41}, y_{51}, y_{61}, y_{12}, y_{22}, y_{32}, y_{42}, y_{52}, y_{62})^T \\
&= (512, 353, 120, 138, 53, 22, 89, 17, 202, 131, 94, 24)^T, \\
&(n_{11}, n_{21}, n_{31}, n_{41}, n_{51}, n_{61}, n_{12}, n_{22}, n_{32}, n_{42}, n_{52}, n_{62})^T \\
&= (825, 560, 325, 417, 191, 373, 108, 25, 593, 375, 393, 341)^T.
\end{aligned}$$

Observe that each row $\boldsymbol{x}_{ij}^T$ of $\boldsymbol{X}$ is also in lexicographical order. In Table 8 the estimated probabilities and diagnostics associated with the $I = 12$ observations are given. The corresponding index plot is in part below of Figure 2 where taking into account the criterion given in Definition 6 and the thresholds shown in Table 9, remarkable observations are highlighted through diamond symbols.

i	j	$\pi(\boldsymbol{x}_{ij}^T\widehat{\boldsymbol{\beta}})$	$r_{ij}^*(\widehat{\boldsymbol{\beta}})$	$h_{ij,ij}(\widehat{\boldsymbol{\beta}})$	$\nu_{ij}(\widehat{\boldsymbol{\beta}})$	$\widehat{p}(ij, 0.05)$	$\widetilde{D}^{(ij)}(\widehat{\boldsymbol{\beta}})$	$D^{(ij)}(\widehat{\boldsymbol{\beta}})$
1	1	0.6415	–4.0273	0.9031	9.3172	0.2017	151.117	207.170
2	1	0.6315	–0.2797	0.9599	23.9321	0.4256	1.873	1.941
3	1	0.3361	1.8808	0.5492	1.2181	0.0004	4.309	4.383
4	1	0.3290	0.1413	0.6580	1.9242	0.0050	0.038	0.039
5	1	0.2392	1.6335	0.4223	0.7311	0.0000	1.951	2.017
6	1	0.0615	–0.3026	0.5358	1.1542	0.0003	0.106	0.102
1	2	0.6642	4.0273	0.2366	0.3100	0.0000	5.028	4.950
2	2	0.6544	0.2797	0.0755	0.0817	0.0000	0.006	0.006
3	2	0.3588	–1.8808	0.7603	3.1724	0.0287	11.223	10.987
4	2	0.3514	–0.1413	0.6317	1.7150	0.0029	0.034	0.034
5	2	0.2578	–1.6335	0.7330	2.7455	0.0187	7.326	6.781
6	2	0.0676	0.3026	0.5345	1.1483	0.0003	0.105	0.109

Table 8: Diagnostics in data from Vickel et al. (1975).

Data	$\overline{\nu}(\widehat{\beta})$ for $\widehat{p}(i,\alpha)$	$2\overline{\nu}(\widehat{\beta})$ for $\nu_i(\widehat{\beta})$	$\overline{\nu}(\widehat{\beta})\chi_{0.05,1}$ for $\widetilde{D}^{(i)}(\widehat{\beta})$ and	$(\chi^2_{0.5,k+1})$ $D^{(i)}(\widehat{\beta})$	k
Finney (1947)	0.0869	0.1738	0.3338	(2.3660)	2
Brown (1980)	0.1440	0.2880	0.5533	(5.3481)	5
Feigl and Zelen (1965)	0.1472	0.2945	0.5657	(2.3660)	2
Vickel at al. (1975)	3.9541	7.9083	15.1898	(6.3458)	6

Table 9: Thresholds for leverage and influence measures.

Conclusions

Inspecting Tables 2, 4, 6 and 8 we can conclude that overall accuracy of the approximation of $D^{(i)}(\widehat{\boldsymbol{\beta}})$ through $\widetilde{D}^{(i)}(\widehat{\boldsymbol{\beta}})$ is very good, only in the most extreme values of $D^{(i)}(\widehat{\boldsymbol{\beta}})$ could be quite bad, in any case it seems that in such extreme cases it holds $\widetilde{D}^{(i)}(\widehat{\boldsymbol{\beta}}) < D^{(i)}(\widehat{\boldsymbol{\beta}})$. For instance, the worst approximation of all the examples is in Feigl and Zelen (1965), where $\widetilde{D}^{(15)}(\widehat{\boldsymbol{\beta}}) = 9.9398 << D^{(15)}(\widehat{\boldsymbol{\beta}}) = 143.6078$. We recommend not using the approximate value $\widetilde{D}^{(i)}(\widehat{\boldsymbol{\beta}})$ in order to be exhaustive using the criterion given in Definition 6, because for example referred to the case 48 in data from Brown (1980), it holds $\widetilde{D}^{(48)}(\widehat{\boldsymbol{\beta}}) = 0.5182 < \overline{\nu}(\widehat{\beta})\chi^2_{0.05,1} = 0.5657 < D^{(48)}(\widehat{\boldsymbol{\beta}}) = 0.6248$ and the conclusion for classifying such a case as influential would be different for both measures.

Although index plots of $D^{(j)}(\widehat{\boldsymbol{\beta}})$ are useful graphical devices, when there are a great amount of observations such plots lose visual effectiveness. Unless there is no any observation that stand out clearly from the rest, it is easy that all $D^{(j)}(\widehat{\boldsymbol{\beta}})$ values seem to be about the same, even some of them are influential. Index plots tend to be conservative. This is actually that happens in Examples 2 and 3. In Zelterman [24, Section 3.3] several influence measures are provided for Example 2 and therein apart from observation 24 another cases are flagged as influential, as it occurs through the criterion given in Definition 6 (see the diamond points referred to the influential analysis through the criterion given in Definition 6 in part below of Figure 1). In Muñoz-García et al. [18] by analyzing Example 3 through two influence power divergence measures it is also concluded that cases 16 and 17 are influential, obviously more moderately than case 15 (see the diamond points referred to the influential analysis through the criterion given in Definition 6 in part above of Figure 2).

To apply the practical operational rule consisting in classify cases with $D^{(j)}(\widehat{\boldsymbol{\beta}})$ values greater than 50-th percentile of the chi-square distribution mentioned in Section 1, in Table 9 the cutoff values are provided (see the column referred to as $\chi^2_{0.5,k+1}$). As it happens for the index plots it seems that such a rule tends to highlight the most influential observations (see Tables 2, 4 and 6 referred to Examples 1, 2 and 3 respectively), actually does not detect any influential case in Example 2. However in Example 4, applying the mentioned rule apart from case (1,1) we should also consider (3,2) and (5,2) as influential (see Table 8), and having so many influential cases in only 12 observations (it affects a quarter part of the data and the middle of departments) does not seem logical. As it was suggested in Muller and Chen Mok [17] focused on linear regression, we do not recommend such a rule.

It is well known that when more than one outlier appears the problem of looking for these observations becomes more difficult due essentially to the masking (false negative outlier detection because the presence of another outlier) and swamping effects (false positive outlier detection because the presence of another outlier). Influential observations may pull the fitted model towards itself which means that its own residual is reduced and the residuals for the rest of observations are increased. However the computational effort for identifying influence cases through the group-deletion instead of the case-deletion is usually quite high, and most of statistical packages include only the simplest one. Moreover sometimes the influence analysis is solved only by the case-deletion. For instance, in Example 4 if we take out the department A, the leverage cases and also outliers vanish, and because $X^2(\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{2}\sum_{j=1}^{6} r^*_{ij}(\widehat{\boldsymbol{\beta}})^2 = 2.5582$, p-value$(X^2(\widehat{\boldsymbol{\beta}})) = P(\chi^2_{(I-k-1)-2} > X^2(\widehat{\boldsymbol{\beta}})) = 0.2783$ with $\widehat{\beta}_1^S = 0.0153$, or $X^2(\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{2}\sum_{j=1}^{6} r^*_{ij}(\widehat{\boldsymbol{\beta}})^2 = 2.6903$, p-value$(X^2(\widehat{\boldsymbol{\beta}})) = P(\chi^2_{(I-k-1)-2+1} > X^2(\widehat{\boldsymbol{\beta}})) = 0.4419$ with $\beta_1^S = 0$, we can conclude that the admission model is conditionally independent to sex.

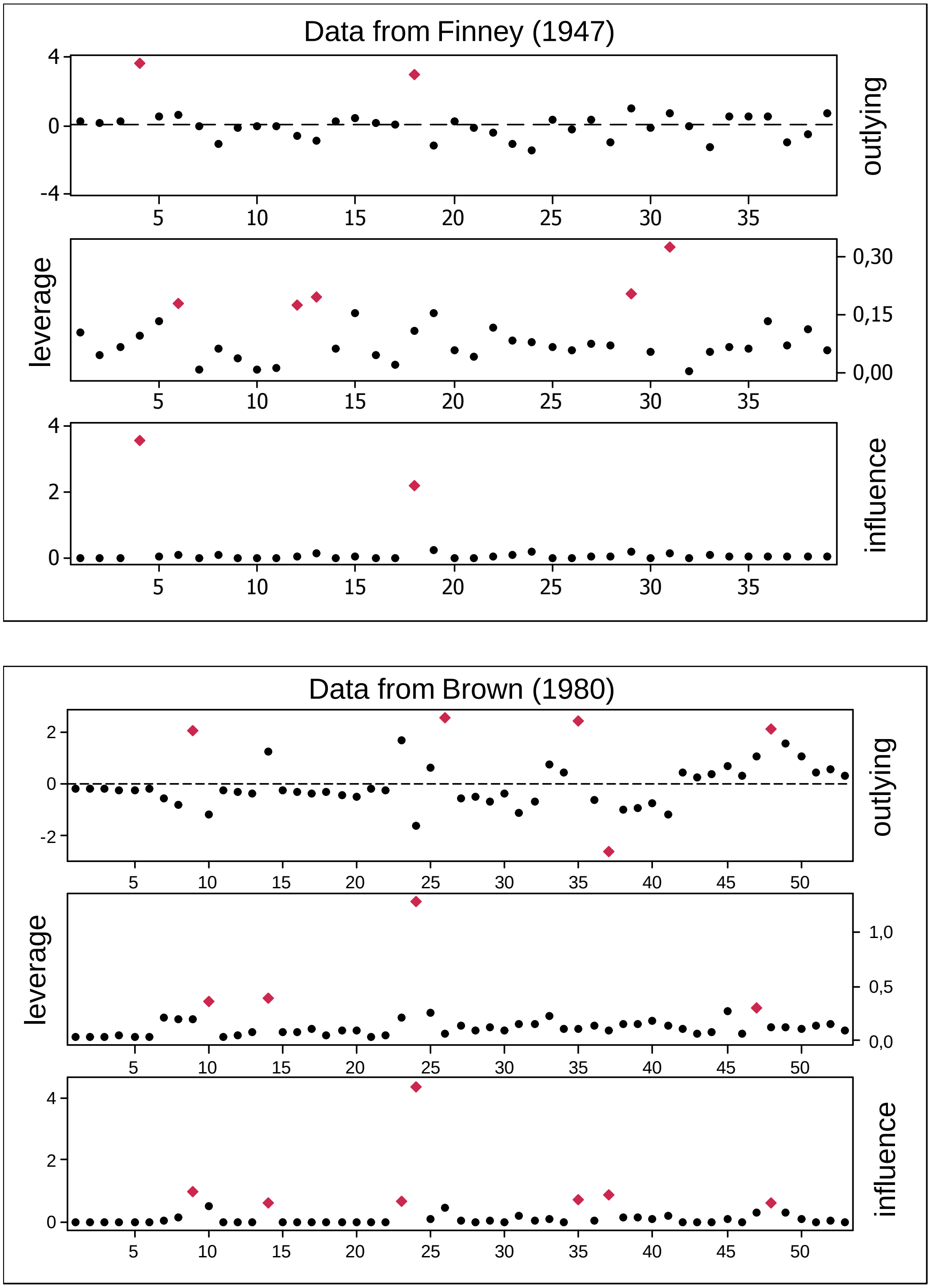
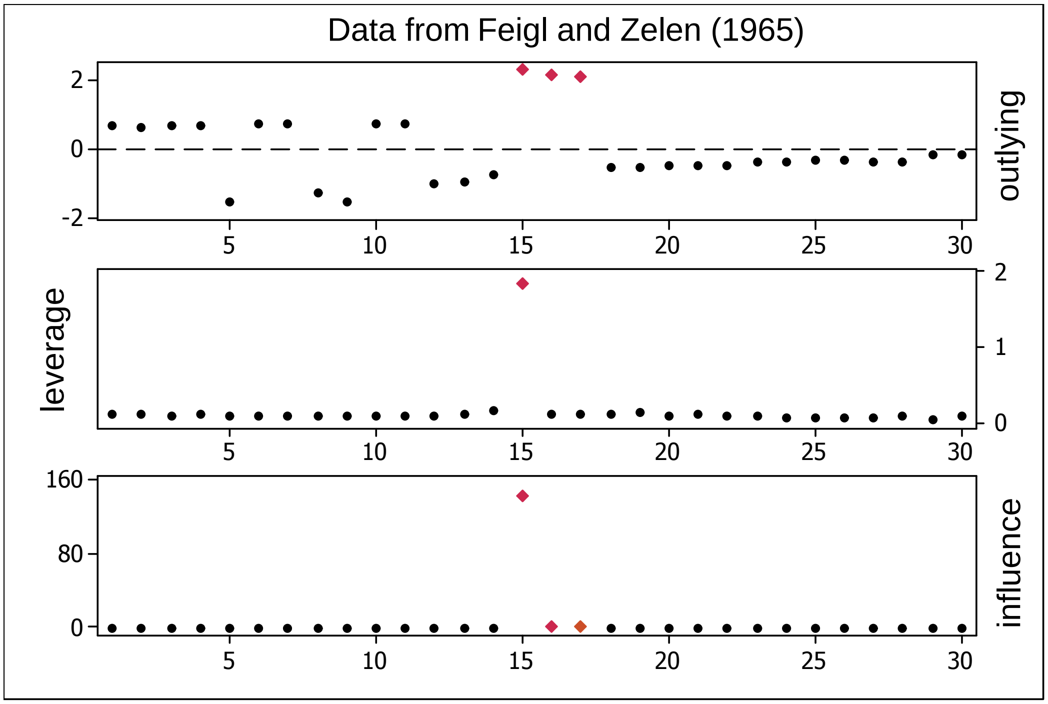


Figure 1: Index Plots for Examples 1 and 2.



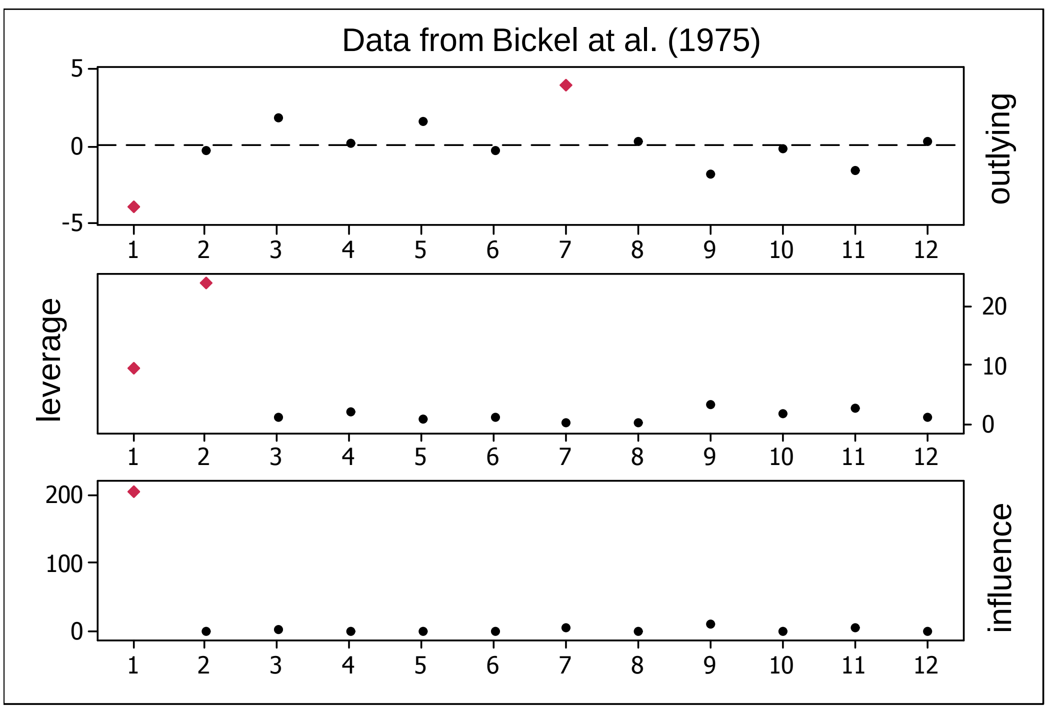


Figure 2: Index Plots for Examples 3 and 4.

References

[1] Agresti, A. (2002). *Categorical Data Analysis* (Second Edition). John Wiley & Sons.

[2] Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley & Sons.

[3] Andersen, E. B. (1992). Diagnostics in categorical data analysis. *Journal of the Royal Statistical Society*, Series B, **54**, 781–791.

[4] Belsley, D. A., Kuh, E. and Welsch, R. E. (2004): *Regression diagnostics: identifying influential data and sources of collinearity* (Second Edition). John Wiley & Sons.

[5] Bickel, P.J., Hammel, E.A., and O'Conner, J.W. (1975). Sex bias in graduate admissions: data from Berkeley. *Science*, **187**, 398–404.

[6] Brown, B. W. (1980). Prediction analysis for binary data. In: *Biostatistics Casebook*, 3-18.

[7] Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, **19**, 15–18.

[8] Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall.

[9] Díaz-García, J.A., González-Farías, G. and Alvarado-Castro, V.M. (2007). Exact distributions for sensivity analysis in linear regression. *Applied Mathematical Sciences*, **1**, 1083–1100.

[10] Dik, J. J. and de Gunst, M. C. M. (1985). The distribution of general quadratic forms in normal variables. *Statistica Neerlandica*, **39**, 14–26.

[11] Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Texts in Statistical Science. Chapman & Hall.

[12] Finney, D. J. (1947). The estimation from individual records of the relationship between dose and quantal response. *Biometrika*, **34**, 320–334.

[13] Feigl, P. and Zelen, M. (1965). Estimation of exponential probabilities with concomitant information. *Biometrics*, **21**, 826–838.

[14] Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression* (Second Edition). John Wiley & Sons.

[15] Jensen, D. R. and Ramirez, D. E. (1998). Some exact properties of Cook's D_I. In *Handbook of Statistics*, Eds. Balakrishnan, N. and Rao, C., Vol. **16**, 387–402, Elsevier Science.

[16] Johnson, W. (1985). Influence measures for logistic regression: Another point of view. *Biometrics,* **72,** 59–65.

[17] Muller, K. E. and Chen Mok, M. (1997). The distribution of Cook's D statistics. *Communications in Statistics - Theory and Methods*, **26**, 525–546.

[18] Muñoz-García, J., Muñoz-Pichardo, J. M. and Pardo, L. (2006). Cressie and Read power-divergences as influence measures for logistic regression models. *Computational Statistics & Data Analysis*, **50**, 3199–3221.

[19] Obenchain, R. L. (1977). Letter to the editor. *Technometrics*, **19**, 348–351.

[20] Pardo, J. A., Pardo, L. and Pardo, M. C. (2005). Minimum ϕ-divergence estimator in logistic regression models. *Statistical Papers, **47**, 91–108.*

[21] Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, **9**, 705–724.

[22] Rao, C. R. and Toutenburg, H. (1999): *Linear Models: Least Squares and Alternatives* (Second Edition). Springer.

[23] Weisberg, S. (1980). *Applied Linear Regression*. John Wiley & Sons.

[24] Zelterman, D. (2005). *Models for Discrete Data.* Oxford University Press.