PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **16** (2004), 35-41

PLISKA
STUDIA MATHEMATICA
BULGARICA

APPLICATION OF THE D-FULLNESS TECHNIQUE FOR BREAKDOWN POINT STUDY OF THE TRIMMED LIKELIHOOD ESTIMATOR TO A GENERALIZED LOGISTIC MODEL

Rositsa Dimova, Neyko Neykov[1]

A new definition for a d–fullness of a set of functions is proposed and its equivalence to the original one given by Vandev [11] is proved. The breakdown point of the WTL_k estimator of Vandev and Neykov [13] for a grouped binary linear regression model with generalized logistic link is studied.

1. Introduction

The classical Maximum Likelihood Estimator (MLE) can be very sensitive to outliers in the data. In fact, even a single outlier can ruin totally the ML estimate. A modification of the MLE, called the Weighted Trimmed Likelihood (WTL) estimator, was proposed by Hadi and Luceño [4], and Vandev and Neykov [13]. Depending on the weights choice and the trimming constant, the WTL estimator reduces to the MLE, to the Least Median of Squares (LMS) and Least Trimmed Squares (LTS) estimators in the normal regression cases, to the Minimum Volume Ellipsoid (MVE) and Minimum Covariance Determinant (MCD) estimators of the multivariate location and scatter in the multivariate normal cases, considered in

[1]Research partially supported by contracts: PRO-ENBIS: GTC1-2001-43031
2000 *Mathematics Subject Classification:* 62J12, 62F35
Key words: Breakdown Point, Subcompact Function, d-fullness, Robustness, Trimmed Likelihood Estimator, generalized logistic model.

details by Rousseeuw and Leroy [10] (see Vandev and Neykov [12],[13]). The Breakdown Point (BP) properties of the WTL estimator were studied by Vandev and Neykov [13], Atanasov and Neykov [1], and Müller and Neykov [7] using the d–fullness technique proposed by Vandev [11]. According to Vandev and Neykov [13], a set $F = \{f_1, \ldots, f_n\}$ of arbitrary functions $f_i : \Theta \to \mathbb{R}^+$, $\Theta \subseteq \mathbb{R}^q$, is called d–full, if for every subset $J \subset \{1, \ldots, n\}$ of cardinality d ($|J| = d$), the function $g_J(\theta) = \max_{j \in J} f_j(\theta)$, $\theta \in \Theta$, is subcompact. A function $g : \Theta \to \mathbb{R}$, $\Theta \subseteq \mathbb{R}^q$, is called subcompact, if its Lebesgue set $L_g(C) = \{\theta \in \Theta : g(\theta) \leq C\}$ is a compact set for every real constant C.

In this paper a new definition for a d–full set of functions, which is equivalent to the original one given by Vandev [11], is proposed. The breakdown point of WLT_k estimator for a grouped binary linear regression model with generalized logistic link is studied.

2. The d-fullness Technique

We remind the replacement variant of the finite sample BP given in Hampel et al. [5], which is closely related to that introduced by Donoho and Huber [3]. Let $X = \{x_i \in \mathbb{R}^p, \text{ for } i = 1, \ldots, n\}$ be a sample of size n.

Definition 1. *The BP of an estimator T at X is given by*

$$\varepsilon_n^*(T) = \frac{1}{n} \max\{m : \sup_{\tilde{X}_m} \|T(\tilde{X}_m)\| < \infty\},$$

where $\tilde{X}_m$ is a sample obtained from X by replacing any m of the points in X by arbitrary values and $\|.\|$ is the Euclidean norm.

We now recall the definition of the Weighted Trimmed estimator given in Vandev and Neykov [13]. Let $F = \{f_i : \Theta \to \mathbb{R}^+, \text{ for } i = 1, \ldots, n\}$ where $\Theta \subseteq \mathbb{R}^q$ is an open set.

Definition 2. *The Weighted Trimmed estimator is defined as*

$$\mathrm{W}_k := \arg\min_{\theta \in \Theta} \sum_{i=1}^{k} w_{\nu(i)} f_{\nu(i)}(\theta), \tag{1}$$

where $f_{\nu(1)}(\theta) \leq f_{\nu(2)}(\theta) \leq \ldots \leq f_{\nu(n)}(\theta)$ are the ordered values of f_i at θ, $\nu = (\nu(1), \ldots, \nu(n))$ is the corresponding permutation of the indices, which depends on θ, k is the trimming parameter, the weights $w_i \geq 0$ for $i = 1, \ldots, n$, are associated with the functions $f_i(\theta)$, and are such that $w_{\nu(k)} > 0$.

Let $x_i \in \mathbb{R}^p$ for $i = 1, \ldots, n$ be i.i.d. observations with probability density function $\phi(x, \theta)$, which depends on an unknown parameter $\theta \in \Theta \subseteq \mathbb{R}^q$ and $f_i(\theta) = -\log \phi(x_i, \theta)$ then the estimator W_k coincides with the WTL_k estimator proposed by Hadi and Luceño [4], and Vandev and Neykov [13].

Vandev and Neykov [13] noted that the set W_k is a nonempty set contained in a compact set if the set $F = \{f_1, \ldots, f_n\}$ is d–full and $k \geq d$, thus there exists a solution of the optimization problem (1).

We need the following notation in the paper. Let f be a function such that $f : \Theta \to \mathbb{R}$, $\partial\Theta$ be the set of the boundary points of Θ, and $\Theta_\infty = \{\{\theta_k\}_{k=1}^\infty : \theta_k \in \Theta, \|\theta_k\| \to \infty\}$ be the set of all sequences whose norm tends to infinity. Then $\underline{f}$ is defined as

$$\underline{f} = \begin{cases} \inf\limits_{\theta^* \in \partial\Theta} \liminf\limits_{\theta_k \to \theta^*} f(\theta_k), & \text{if } \Theta \text{ is bounded, or} \\[2ex] \inf\limits_{\theta^* \in \partial\Theta} \liminf\limits_{\substack{\theta_k \to \theta^* \\ \{\theta_k\} \in \Theta_\infty}} f(\theta_k), & \text{if } \Theta \text{ is unbounded.} \end{cases} \tag{2}$$

Proposition 1. *A continuous function f is a subcompact if and only if $\underline{f} = \infty$.*

To prove this, we will use the following two lemmas.

Lemma 1. *(Demidenko [2]) Let $f : \Theta \to R$ be continuous function, $\Theta \subseteq R^q$ be an open set, and there exists $\theta_0 \in \Theta$, such that $f(\theta_0) < \underline{f}$. Then the set $S_0 = \{\theta : f(\theta) \leq f(\theta_0)\}$ is a nonempty compact set.*

Lemma 2. *Let $f : \Theta \to R$ be continuous function, $\Theta \subseteq R^q$ be an open set. If there exists $\theta_0 \in \Theta$, such that the set $S_0 = \{\theta : f(\theta) \leq f(\theta_0)\}$ is a compact set, then $f(\theta_0) < \underline{f}$.*

Proof. Let $\{\theta_k\}_{k=1}^\infty$ be a sequence from Θ such that $\theta_k \to \theta^*$, $\theta^* \in \partial\Theta$. Hence $\theta^* \notin S_0$ (since Θ is an open set and $S_0 \subset \Theta$) and $\theta_k \notin S_0$ (since S_0 is a compact set), that is $f(\theta_k) > f(\theta_0)$. The function f is continuous, therefore $\liminf\limits_{\theta_k \to \theta^*} f(\theta_k) > f(\theta_0)$.

Now let $\{\tilde{\theta}_k\}_{k=1}^\infty$ be a sequence from Θ such that $\|\tilde{\theta}_k\| \to \infty$. Hence $\tilde{\theta}_k \notin S_0$ (S_0 is a compact set), that is $f(\tilde{\theta}_k) > f(\theta_0)$. Since f is continuous, we have that $\liminf\limits_{\|\tilde{\theta}_k\| \to \infty} f(\tilde{\theta}_k) > f(\theta_0)$.

Therefore $\underline{f} > f(\theta_0)$ as the sequences $\{\theta_k\}_{k=1}^\infty$ and $\{\tilde{\theta}_k\}_{k=1}^\infty$ are arbitrary. $\square$

Proof of Proposition 1: Let f be a subcompact function, and θ_0 be an arbitrary point from Θ, therefore the set $S_0 = \{\theta : f(\theta) \leq f(\theta_0)\}$ is a compact set. Using Lemma 2 we obtain that $f(\theta_0) < \underline{f}$, but θ_0 is arbitrary, that is $\underline{f} = \infty$.

Now let $\underline{f} = \infty$. Hence $f(\theta_0) < \underline{f}$ for every $\theta_0 \in \Theta$. From Lemma 1 follows that the set $S_0 = \{\theta : f(\theta) \leq f(\theta_0)\}$ is a compact set, that is the function f is a subcompact function.

Using the above proposition, we obtain the equivalent definition for a d–full set of functions, as follows.

Definition 3. *A set $F = \{f_1, \ldots, f_n\}$ of arbitrary functions $f_i : \Theta \to \mathbb{R}$, $\Theta \subseteq \mathbb{R}^q$, is called d–full if $\underline{g_{_J}} = \infty$ for every subset $J \subset \{1, \ldots, n\}$ of cardinality d, where $g_{_J}(\theta) = \max_{j \in J} f_j(\theta)$.*

For some classes of probability distributions, e.g., the class of normal mixtures distributions, the corresponding set $F = \{-\log \phi(x_i, \theta), i = 1, \ldots, n\}$ is not d-full for any d since $\underline{g_{_J}} < \infty$.

In the next section we will study the BP of a grouped binary linear regression model with generalized logistic link.

3. Grouped binary linear regression model with generalized logistic link

The type of the data under consideration is of the form (y_i, x_i^T) for $i = 1, \ldots, m$. It is assumed that, y_i is binomially distributed, $b(y_i \mid n_i, \pi_i)$, where the group size is n_i, the probability of success is π_i, and x_i is a p–dimensional vector of covariates (explanatory variables). The total number of observations is $n = n_1 + n_2 + \cdots + n_m$. We will assume that $0 < y_i < n_i$ for each i, and π_i follows the Prentice [8] generalized logistic distribution

$$\pi_i = (1 + \exp(-\eta_i))^{-a},$$

where $a > 0$, $\eta_i = x_i^T \beta$ is the linear predictor and β is a p–dimensional vector of unknown parameters.

The particular case, when a=1, is considered by Müller and Neykov [7] who proved that the BP of the WTL$_k$ estimator is equal to $\min(m-k+1, k-\mathcal{N}(X))/m$, where $\mathcal{N}(X) = \max_{0 \neq \beta \in \mathbb{R}^p} \operatorname{card}\{i \in \{1, \ldots, m\}; x_i^T \beta = 0\}$.

We will show that the set $F = \{f(y_i, \eta_i, a), i = 1, \ldots, m\}$, where $f(y_i, \eta_i, a) = -\log \binom{n_i}{y_i} + y_i a \log(1 + e^{-\eta_i}) - (n_i - y_i) \log(1 - (1 + e^{-\eta_i})^{-a})$, is d–full following Definition 3.

It is obvious that $\lim\limits_{a \to 0} f(y_i, \eta_i, a) = +\infty$, $\lim\limits_{a \to +\infty} f(y_i, \eta_i, a) = +\infty$, and $\lim\limits_{\eta_i \to \pm\infty} f(y_i, \eta_i, a) = +\infty$. Therefore $f(y_i, \eta_i, a)$ is a subcompact function because $\underline{f} = +\infty$.

Proposition 2. *The set* $\{f(y_i, x_i, \beta, a), i = 1, \ldots, m\}$ *is* $\mathcal{N}(X) + 1$*-full.*

P r o o f. Let $C \in \mathbb{R}$ is arbitrary. Since $f(y_i, \eta_i, a)$ is a subcompact function of η_i and a, there exist constants B_i and A_i for $i \in I \subset \{1, \ldots, m\}$, $\mathrm{card}(I) = \mathcal{N}(X)+1$, such that the set:

$$\begin{aligned}
&\{\beta \in \mathbb{R}^p, a > 0 : \max_{i \in I} f(y_i, x_i, \beta, a) \le C\} \\
= \; & \bigcap_{i \in I} \{\beta \in \mathbb{R}^p, a > 0 : f(y_i, x_i, \beta, a) \le C\} \\
= \; & \bigcap_{i \in I} \{\beta \in \mathbb{R}^p, a > 0 : f(y_i, \eta_i = x_i^T \beta, a) \le C\} \\
\subset \; & \bigcap_{i \in I} \{\{\beta \in \mathbb{R}^p : |x_i^T \beta| \le B_i\} \times \{a : 0 < a \le A_i\}\}
\end{aligned}$$

is contained in a compact set. (The set $\{\beta \in \mathbb{R}^p | x_i^T \beta| \le B_i\}$ is bounded for all B_i according to Lemma 3 of Müller and Neykov [7].) □

As a consequence of this proposition, the following corollary is obtained.

Corollary 1. *The set* W_k *for the grouped binary linear regression model with generalized logistic link is a non empty compact set if* $k \ge \mathcal{N}(X) + 1$.

Applying Theorem 2 of Müller and Neykov [7], which states that if the set $F = \{-\log \phi(x_i, \theta), i = 1, \ldots, n\}$ (here $x_i, i = 1, \ldots, n$ are i.i.d. observations with p.d.f. $\phi(x, \theta)$) is d-full, $\lfloor (n + d)/2 \rfloor \le k \le \lfloor (n + d + 1)/2 \rfloor$, then the BP of the WTL estimator satisfies $\varepsilon_n^*(W_k) \ge \frac{1}{n} \left\lfloor \frac{n - d + 2}{2} \right\rfloor$, we get the following

Corollary 2. *The BP of the* WTL_k *estimator for the grouped binary linear regression model with generalized logistic link is*

$$\varepsilon_m^*(W_k) \ge \frac{1}{m} \left\lfloor \frac{m - \mathcal{N}(X) + 1}{2} \right\rfloor$$

if $\lfloor (m + \mathcal{N}(X) + 1)/2 \rfloor \le k \le \lfloor (m + \mathcal{N}(X) + 2)/2 \rfloor$.

We remind that $\lfloor z \rfloor := \max\{n : n \le z\}$.

REFERENCES

[1] D. V. Atanasov, N. M. Neykov. On the finite sample breakdown point of the weighted trimmed likelihood estimators and the d–fullness of a set of continuous functions. In: *Proc. of the VI Intern. Conf. "Computer Data Analysis and Modelling"* **1** (2001), 52–57.

[2] E.Z. Demidenko. Optimization and Regression, Nauka, Moscow, 1989 (in Russian).

[3] D.L. Donoho, P.J. Huber. The notion of breakdown point. *A festschrift for Eric Lehmann.* Belmont, CA: Wadsworth, (1983), 157–184.

[4] A.S. Hadi, A. Luceño. Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics and Data Analysis.* **25** (1997), 251-272.

[5] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W. A. Stahel. Robust statistics. The approach based on influence functions. John Wiley & Sons, Inc, 1986.

[6] P. Huber. Robust Statistics. John Wiley & Sons, New York, 1981.

[7] Ch.H. Müller, N.M. Neykov. Breakdown points of the trimmed likelihood and related estimators in generalized linear models. *J. Statist. Plann. Inference.* **116** (2003), 503–519.

[8] R.L. Prentice. A generalization of the probit and logit methods for dose response curves. *Biometrics.* **32** (1976), 761–768.

[9] P.J. Rousseeuw. Multivariate Estimation with High Breakdown Point. *Mathematical Statistics and Applications, Reidel Publishing Company* **I** (1986), 283–297.

[10] P.J. Rousseeuw, A. Leroy. Robust regression and outlier detection. John Wiley & Sons, New York, 1987.

[11] D.L. Vandev. A note on breakdown point of the least median squares and least trimmed squares. *Statistics and Probability Letters* **16** (1993), 117–119.

[12] D.L. Vandev, N.M. Neykov. Robust maximum likelihood in the Gaussian case. *New Directions in Statistical Data Analysis and Robustness*, Basel, Birkhauser Verlag, 1993, 257-264.

[13] D.L. Vandev, N.M. Neykov. About regression estimators with high breakdown point. *Statistics* **32** (1998), 111–129.

Rositsa Dimova
Faculty of Mathematics and Informatics
Sofia University "St. KlimentOhridski"
5 J. Bourchier Blvd., 1164 Sofia, Bulgaria
email: `rdimova@fmi.uni-sofia.bg`

Neyko Neykov
National Institute of Meteorology and Hydrology
Bulgarian Academy of Sciences
66 Tsarigradsko chaussee, 1784 Sofia, Bulgaria
email: `neyko.neykov@meteo.bg`