Trimmed likelihood estimators in generalized linear models

Christine H. Müller
Carl von Ossietzky University, Department of Mathematics
Postfach 2503, D - 26111 Oldenburg, GERMANY
e-mail: `mueller@math.uni-oldenburg.de`

Abstract: Trimmed likelihood estimators in linear models are the least trimmed squares estimators and it is known that they have high breakdown points. In particular their breakdown points depend only on the trimming proportion and a quantity $\mathcal{N}(X)$ introduced by Müller (1995). Thus, the trimming proportion which maximizes the breakdown point of trimmed likelihood estimators in linear models depends on $\mathcal{N}(X)$ as well. Now it has turned out, that the same dependence holds also for trimmed likelihood estimators in generalized linear models. The proof of this result uses the concept of the fullness parameter of Vandev (1993) and Vandev and Neykov (1998). For calculating trimmed likelihood estimators, we propose a genetic algorithm.

1 Introduction

Assume that the distribution of the observations Y_n have densities $f(y_n, x_n, \beta)$ given by a linear exponential family, that is

$$f(y_n, x_n, \beta) = \exp\{T(y_n)^\top g(x_n^\top \beta) + c(x_n^\top \beta) + h(y_n)\},$$

where $T : \mathcal{Y} \to \Re^r$, $g : \Re \to \Re^r$, $c : \Re \to \Re$, and $h : \mathcal{Y} \to \Re$ are known functions with $\mathcal{Y} \subset \Re^q$, $x_n \in \mathcal{X} \subset \Re^p$, $n = 1, \ldots, N$, are known explanatory variables and $\beta \in \Re^p$ is unknown. The observations $Y_1, \ldots, Y_N$ are independent. Let $y := (y_1, \ldots, y_N)^\top$ the vector of all realized observations, $X := (x_1, \ldots, x_N)^\top$ the design matrix,

$$\begin{aligned} l_n(y, X, \beta) &:= -\log f_n(y_n, x_n, \beta) \\ &= -T(y_n)^\top g(x_n^\top \beta) - c(x_n^\top \beta) - h(y_n), \end{aligned}$$

the log-likelihood function. Maximum likelihood (ML) estimators are maximizing the likelihood, i.e. minimzing $\sum_{n=1}^{N} l_n(y, X, \beta)$. Trimming the least likely observations, i.e. the observations with the largest $l_n(y, X, \beta)$, leads to trimmed likelihoods. Maximizing the trimmed likelihood provides the trimmed likelihood estimators $TL_h(y)$ given by

$$TL_h(y) := \arg\min_{\beta} \sum_{n=1}^{h} l_{(n)}(y, X, \beta),$$

where $N - h$ observations are trimmed and $l_{(1)}(y, X, \beta) \leq \ldots \leq l_{(N)}(y, X, \beta)$ are the ordered log-likelihoods.

In the case of normal distribution with known variance, the trimmed likelihood estimators coincide with the least trimmed squares (LTS) estimators of Rousseeuw (1984, 1985) and Rousseeuw and Leroy (1987). Breakdown points of LTS estimators for linear regression were derived in Rousseeuw (1984, 1985), Rousseeuw and Leroy (1987), Vandev (1993), Vandev and Neykov (1993), Coakley and Mili (1993), Hössjer (1994), Müller (1995, 1997), Mili and Coakley (1996); and Hössjer (1994) showed also consistency and asymptotic normality. Trimmed likelihood estimators for normal distribution with unknown variance were regarded in Bednarski and Clarke (1993) who derived their asymptotic properties like Fisher consistency, asymptotic normality and compact differentiability.

Up to now, not much is known about trimmed likelihood estimators for distributions different from the normal distribution. There are approaches on robust and in particular high breakdown point estimators for logistic regression and other generalized linear models given by Stefanski, Carroll, and Ruppert (1986), Copas (1988), Künsch, Stefanski and Carroll (1989), Carroll and Pederson (1993), Wang and Carroll (1993, 1995), Christmann (1994), Hubert (1997), Christmann and Rousseeuw (1999). But these approaches do not concern trimmed likelihood estimators.

Only Vandev and Neykov (1998) derived breakdown points of trimmed likelihood estimators for logistic regression and exponential linear models with unknown dispersion. Their approach bases on the concept of d-fullness developed by Vandev (1993). However, they could only derive breakdown points under the restriction that the explanatory variables $x_1, \ldots, x_N$ of the logistic regression and the exponential linear model are in general position. This restriction was also used in the approaches of Rousseeuw (1984, 1985) and Rousseeuw and Leroy (1987) concerning LTS estimators in linear models. Müller (1995, 1997) and Mili and Coakley (1996) dropped this restriction and showed that then the breakdown point of LTS estimators is determined by $\mathcal{N}(X)$ defined as

$$\mathcal{N}(X) \quad := \quad \max_{0 \neq \beta \in \Re^p} \text{card} \left\{ n \in \{1, \ldots, N\};\ x_n^\top \beta = 0 \right\}.$$

If the explanatory variables are in general position then $\mathcal{N}(X) = p - 1$ which is the minimum value for $\mathcal{N}(X)$. In other cases $\mathcal{N}(X)$ is much higher. These other cases appear mainly when the explanatory variables are qualitative as in ANOVA models or given by an experimenter in a designed experiment.

In Section 2 of this paper, we show that the quantity $\mathcal{N}(X)$ determines the breakdown point not only of LTS estimators in linear models but also of any trimmed likelihood estimator in generalized linear models. In particular, we will show how the fullness parameter of Vandev (1993) is connected with $\mathcal{N}(X)$. Section 3 deals with the computation of trimmed likelihood estimators.

2 Breakdown points

The breakdown point of an estimator $\hat{\beta} : \mathcal{Y}^N \to \Re^p$ for β at the sample $y \in \mathcal{Y}^N$ is defined as (compare e.g. Donoho and Huber 1983, Hampel et al. 1986, p. 97)

$$\epsilon^*(\hat{\beta}, y) := \tfrac{1}{N} \min \Big\{M; \;\text{ there exists no bounded set } B \text{ with } \{\hat{\beta}(\overline{y});\ \overline{y} \in \mathcal{Y}_M(y)\} \subset B\Big\},$$

where

$$\mathcal{Y}_M(y) := \left\{\overline{y} \in \mathcal{Y}^N;\ \text{card}\{n;\ y_n \neq \overline{y}_n\} \leq M\right\}.$$

Vandev and Neykov (1998) connected the notion of breakdown point of trimmed likelihood estimators with the notion of d-fullness which is based on the concept of sub-compact functions. If the parameter space is $\Re^p$, then every compact set is included in a bounded set and vice versa. Hence, in the case of estimating $\beta \in \Re^p$, we can use boundeness instead of compactness so that we have the following specifications of the definitions of Vandev and Neykov (see also Müller and Neykov 2000).

Definition 1 *A function $\gamma : \Re^p \to \Re$ is called sub-bounded if the set $\{\beta \in \Re^p;\ \gamma(\beta) \leq C\}$ is bounded for all $C \in \Re$.*

Definition 2 *A finite set $\Gamma = \{\gamma_n : \Re^p \to \Re;\ n = 1, \ldots, N\}$ of functions is called d-full if for every $\{n_1, \ldots, n_d\} \subset \{1, \ldots, N\}$ the function γ given by $\gamma(\beta) := \max\{\gamma_{n_k}(\beta);\ k = 1, \ldots, d\}$ is sub-bounded.*

The following theorem presents the connection between the breakdown point of trimmed likelihood estimators and the d-fullness parameter.

Theorem 1 *If $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is d-full, then*

$$\epsilon^*(TL_h, y) \geq \frac{1}{N} \min\{N - h + 1, h - d + 1\}.$$

This theorem holds for general models and not only for generalized linear models (see Müller and Neykov 2000). However, for generalized linear model the d-fullness parameter can be given more explicitly by the quantity $\mathcal{N}(X)$.

Theorem 2 *If the function γ_z given by $\gamma_z(\theta) = -T(z)^\top g(\theta) - c(\theta) - h(z)$ is sub-bounded for all $z \in \mathcal{Y}$ then the family $\{l_n(y, X, \cdot);$ $n = 1, \ldots, N\}$ is $\mathcal{N}(X)+1$-full for all $y \in \mathcal{Y}^N$ and all $X \in \mathcal{X}^N$.*

For logistic and log-linear regression models it can be easily seen that the function γ_z is sub-bounded. Setting $y = (s, t)$ with $t = (t_1, \ldots, t_N)^\top$ and $s = (s_1, \ldots, s_N)^\top$, we have for logistic regression

$$\begin{aligned} & l_n(y, X, \beta) = l_n(s, t, X, \beta) \\ & = \quad -s_n\, x_n^\top \beta + t_n \, \log(1 + \exp(x_n^\top \beta)) - \log \left(\begin{pmatrix} t_n \\ s_n \end{pmatrix} \right) \\ & = \quad \gamma_{s_n, t_n}(x_n^\top \beta). \end{aligned}$$

The function $\gamma_{u,v}$ given by $\gamma_{u,v}(\theta) = -u\theta + v\, \log(1 + \exp(\theta)) - \log \left(\begin{pmatrix} v \\ u \end{pmatrix} \right)$ is sub-bounded as soon as $0 < u < v$ so that the set $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is $\mathcal{N}(X)+1$-full if $y \in \mathcal{Y} := \{y = (s, t);\ 0 < s_n < t_n$ for $n = 1, \ldots, N\}$. The log-linear regression model is given by

$$l_n(y, X, \beta) = -y_n\, x_n^\top \beta + \exp(x_n^\top \beta) + \log(y_n!).$$

The function γ_z given by $\gamma_z(\theta) = -z\theta + \exp(\theta) + \log(z!)$ is sub-bounded as soon as $z > 0$ so that the set $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is $\mathcal{N}(X) + 1$-full for all $y \in \mathcal{Y} = \{y \in \Re^N;\ y_n > 0$ for all $n = 1, \ldots, N\}$.

Since for logistic regression and log-linear models a lower bound for the breakdown point can proved as well (see Müller and Neykov 2000) we have for both models the following result.

Theorem 3 *The breakdown point of a trimmed likelihood estimator TL_h for a logistic regression model or a log-linear model satisfies*

$$\min_{y\in\mathcal{Y}} \epsilon^*(TL_h, y, X) = \frac{1}{N} \min\{N-h+1, h-\mathcal{N}(X)\}.$$

The maximum breakdown point is attained for for h satisfying $\left\lfloor \frac{N+\mathcal{N}(X)+1}{2} \right\rfloor \le h \le \left\lfloor \frac{N+\mathcal{N}(X)+2}{2} \right\rfloor$ and equals $\frac{1}{N} \left\lfloor \frac{N-\mathcal{N}(X)+1}{2} \right\rfloor$.

Theorem 3 shows that the maximum breakdown point value for logistic regression and for log-linear models is the same as for linear models. Also the optimal trimming proportion h coincides. See Müller (1995, 1997).

3 Calculation of the estimators

Trimmed likelihood estimators can be calculated as the LTS estimators for linear models. We implemented two algorithms. One, useful for small data sets, is an exact algorithm and is based on the idea of Agulló (1996) using a branch and bound algorithm. If the data set is large, then a genetic algorithm like that of Burns (1992) used in S-PLUS is only applicable. Our genetic algorithm is based on an old population of ω subsets with h elements and a new population of ν randomly generated subsets. Recombination of all subsets within the old population and recombination of the subsets of the old population with those of the new population provides a population of $\frac{\omega(\omega-1)}{2} + \omega\nu + \omega + \nu$ subsets. From this population the ω best subsets (that with smallest log-likelihood) are chosen as old population for the next step. We applied this algorithm with $\omega = 7 = \nu$ to the eggs data of O'Hara Hines and Carter (1993, p.13) with $N = 46$ observations (observations 32 and 40 were dropped because of zero response). A logistic regression model given by

$$\begin{aligned} \mathrm{logit}(p/(1-p)) &= \beta_1 + \beta_2 * WH + \beta_3 * \log_{10}(Concentration) \\ &+ \beta_4 * \log_{10}(Concentration)^2 \end{aligned}$$

was used. Then the breakdown point maximizing trimming proportion is $h = 36$. We obtained $TL_{36}(y, X) = (7.36, -0.12, -9.29, 2.16)^\top$ as trimmed likelihood estimator for $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^\top$, and the trimmed observations were 13, 14, 20, 21, 38, 39, 41, 42, 43, 44 (see Müller and Neykov 2000).

References

1. Agulló, J. (1996). Exact iterative computation of the multivariate minimum volume ellipsoid estimator with a branch and bound algorithm. In *COMPSTAT'96, Proceedings in Computational Statistics,* ed. A. Prat, Physica-Verlag, 175-180.

2. Bednarski, T. and Clarke, B. R. (1993). Trimmed likelihood estimation of location and scale of the normal distribution. *Austral. J. Statist.* **35**, 141-153.

3. Burns, P.J. (1992). A genetic algorithm for robust regression estimation. *StatSci Technical Note.*

4. Carroll, R.J. and Pederson, S. (1993). On robustness in the logistic regression model. *J. R. Statist. Soc. B* **55**, 693-706.

5. Christmann, A. (1994). Least median of weighted squares in logistic regression with large strata. *Biometrika* **81**, 413-417.

6. Christmann, A. and Rousseeuw, P.J. (1999). Measuring overlap in logistic regression. *Technical Report, University of Antwerp, submitted.*

7. Coakley, C.W. and Mili, L. (1993). Exact fit points under simple regression with replication. *Statist. Probab. Lett.* **17**, 265-271.

8. Copas, J. B. (1988). Binary regression models for contaminated data (with discussion). *J. R. Statist. Soc. B* **50**, 225-265.

9. Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann,* Ed. P.J. Bickel, K.A. Doksum and J.L. Hodges, Jr., pp. 157-84. Belmont, CA: Wadsworth.

10. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics. The Approach Based on Influence Functions.* Wiley, New York.

11. Hössjer, O. (1994). Rank-based estimates in the linear model with high breakdown point. *J. Amer. Statist. Assoc.* **89**, 149-158.

12. Hubert, M. (1997). The breakdown value of the L_1 estimator in contingency tables. *Statistics and Probability Letters.* **33**, 419-425.

13. Künsch, H. R., Stefanski, L. A. and Carroll, R. J. (1989). Conditionally unbiased bounded influence estimation in general regression models, with applications to generalized linear models. *J. Amer. Statist. Assoc.* **84**, 460-466.

14. Mili, L. and Coakley, C. W. (1996). Robust estimation in structured linear regression. *Ann. Statist.* **15**, 2593-2607.

15. Müller, Ch. H. (1995). Breakdown points for designed experiments. *J. Statist. Plann. Inference.* **45**, 413-427.

16. Müller, Ch. H. (1997). *Robust Planning and Analysis of Experiments. Lecture Notes in Statistics.* **124**, Springer, New York.

17. Müller, Ch. H. and Neykov, N. (2000). Breakdown points of trimmed likelihood estimators in generalized linear models. *Submitted.*

18. O'Hara Hines, R.J. and Carter, E.M. (1993). Improved added variable and partial residual plots for the detection of influential observations in generalized linear models. *Appl. Statist.* **42**, 3-20.

19. Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79**, 851-857.

20. Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Vol. B*, eds. W. Grossman, G. Pflug, I. Vincze and W. Wertz. Reidel, Dordrecht, 283-297.

21. Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

22. Stefanski, L.A., Carroll, R.J. and Ruppert, D. (1986). Optimally bounded score functions for generalized linear models with applications to logistic regression. *Biometrika* **73**, 413-424.

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

24. Vandev, D. and Neykov, N.(1998). About regression estimators with high breakdown point. *Statistics.* **32**, 111-129.

25. Wang, C.Y. and Carroll, R.J. (1993). On robust estimation in logistic case-control studies. *Biometrika* **80**, 237-241.

26. Wang, C.Y. and Carroll, R.J. (1995). On robust logistic case-control studies with response-dependent weights. *J. Statist. Plann. Inference* **43**, 331-340.