Journal of Statistical Planning and Inference 151-152 (2014) 60–72

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Iterative rank estimation for generalized linear models

Guy-vanie M. Miakonkana [a,*], Asheber Abebe [b]

[a] *Institute for Mathematics and its Applications, University of Minnesota, 207 Church Street SE, 306 Lind Hall, Minneapolis, MN 55455, United States*

[b] *Auburn University, United States*

ARTICLE INFO

Article history:
Received 30 October 2013
Received in revised form
21 January 2014
Accepted 4 April 2014
Available online 18 April 2014

MSC:
primary 62J12, 62G05;
secondary 62F12, 62G35

Keywords:
Generalized linear model
Wilcoxon
Robustness
Quasi-likelihood

ABSTRACT

In this paper, the estimation of parameters of a generalized linear regression model is considered. The proposed estimator is defined iteratively starting from an initial obtained by minimizing the Wilcoxon dispersion function for independent errors. It is shown that the iterative estimator converges to the rank version of the maximum quasi-likelihood estimator as the number of iterations increases. The consistency and the asymptotic normality of the rank version of the maximum quasi-likelihood estimator are given. As in the linear model, the procedure results in estimators that are robust in the response space. This is proven theoretically via the influence function. A simulation study and a real world data example illustrate the robustness in the response space and efficiency of the estimator.

1. Introduction

Let y be a response variable and $\mathbf{x}$ be a vector of predictors. Assume that both $\mathbf{x}$ and y are random and that we have a random sample $(\mathbf{x}_i, y_i),\ i=1,\dots,n$. We consider the generalized linear regression model

$$h[E(y_i|\mathbf{x}_i=\mathbf{x})]=\mathbf{x}^t\boldsymbol{\theta}_0, \qquad (1.1)$$

where $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ is an unknown true parameter vector. We assume that $y_1, y_2, \dots, y_n$ are independent absolutely continuous random variables with distribution in the exponential family of distributions, $\mathbf{x}_i \in \mathbb{X} \subset \mathbb{R}^p,\ 1 \le i \le n$, are independent random vectors. The function h is a known function such that its inverse $g \equiv h^{-1}$ is a real valued function defined on the set $U=\{u|u=\mathbf{x}^t\boldsymbol{\theta} \text{ for } \boldsymbol{\theta}\in\boldsymbol{\Theta} \text{ and } \mathbf{x}\in\mathbb{X}\}\subset\mathbb{R}$, is monotone and three times continuously differentiable. We shall assume that, $\mathbb{X}$ is compact, $\boldsymbol{\Theta}$ is convex and compact, and $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.

Since the fundamental work of Nelder and Wedderburn (1972), there has been continued interest in the development of the theory and the methodology related to generalized linear models to estimate the parameter $\boldsymbol{\theta}_0$. Wedderburn (1974) proposed the least squares and the quasi-likelihood estimators. These estimators are asymptotically efficient in the sense that the limiting variance–covariance matrix attains a Cramer–Rao-type lower bound. We also refer to McCullagh and Nelder (1989) for a comprehensive account of the generalized linear models and quasi-likelihood based inference procedures. However, in the presence of outliers it is desirable to use a robust estimation procedure. Pregibon (1982),

* Corresponding author.
E-mail address: gmiakonk@umn.edu (G.-v. Miakonkana).

http://dx.doi.org/10.1016/j.jspi.2014.04.002

Stefanski et al. (1986), and Künsch et al. (1989) considered the robust estimation of generalized linear models parameters with particular emphasis on logistic regression. Morgenthaler (1992) studied least absolute deviations fits for generalized linear models. Robust M-estimators in logistic regression model were proposed by Kordzakhia et al. (2001). An alternative way to develop robust estimators is to use rank based procedures. The rank-based approach has not been described in the literature and is the focus of this paper. Our rank based method uses the so-called Wilcoxon objective function that provides us with an initial estimator which is, afterwards, updated iteratively. The procedure results in estimators that are robust in the response space. Extensions of this method give us bounded influence and high-breakdown estimators.

Rank estimation for the linear regression model was proposed by Jurečková (1971) and Jaeckel (1972). Over the years, several extensions and refinements of the rank regression approach were proposed. A comprehensive treatise is given in Hettmansperger and McKean (1998). Particularly relevant to our discussion are the works of Jung and Ying (2003) and Abebe and McKean (2007). Jung and Ying (2003) studied the linear model with correlated and non-i.i.d. errors. Although our proposed method is developed for independent errors, it allows for arbitrary link functions with minimal assumption on the error distribution. Abebe and McKean (2007) studied the Wilcoxon estimator of a general nonlinear regression function. The initial estimator we use in the iteratively defined rank estimator proposed in this paper is a special case of the estimator developed in Abebe and McKean (2007).

The remainder of our paper is organized as follows. After introducing the model and the estimators in Section 2, the consistency and the asymptotic normality of the rank version of the maximum quasi-likelihood estimator are studied in Section 3. We illustrate the robustness and the efficiency of the estimators in Section 4 via simulation studies and a real world data example. Section 5 provides the conclusion. Proofs and technical details are found in Appendix A.

2. Iterative rank estimator

Take $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and define the Pearson residuals as $z_i(\boldsymbol{\theta},\boldsymbol{\theta}_0) = (y_i - g(\mathbf{x}_i^t\boldsymbol{\theta}))/\phi\sqrt{\nu(g(\mathbf{x}_i^t\boldsymbol{\theta}_0))}$, $1 \le i \le n$, where ν is a continuous function (McCullagh and Nelder, 1989, p. 30) related to the variance of y_i through $var(y_i) = \phi^2\nu(g(\mathbf{x}_i^t\boldsymbol{\theta}_0))$, the dispersion parameter $\phi > 0$ is assumed to be an unknown constant and the function ν will be assumed to be twice continuously differentiable. The assumptions on g and ν are the same as $M3$ in Chiou and Müller (1999). Consider the estimator of $\boldsymbol{\theta}$ defined as the minimizer of the rank dispersion function of Jaeckel (1972)

$$W_n(\boldsymbol{\theta},\boldsymbol{\theta}_0) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{R(z_i(\boldsymbol{\theta},\boldsymbol{\theta}_0))}{n+1} - \frac{1}{2}\right] z_i(\boldsymbol{\theta},\boldsymbol{\theta}_0),$$

where $R(z_i(\boldsymbol{\theta},\boldsymbol{\theta}_0))$ is the rank of $z_i(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ among $z_1(\boldsymbol{\theta},\boldsymbol{\theta}_0),\ldots,z_n(\boldsymbol{\theta},\boldsymbol{\theta}_0)$. Since $\boldsymbol{\theta}_0$ is not known, the dispersion function cannot be directly minimized. As a solution, we plug-in an initial estimator of $\boldsymbol{\theta}_0$ in W_n and minimize the resulting dispersion function with respect to $\boldsymbol{\theta}$. That is, given an initial estimator $\widehat{\boldsymbol{\theta}}_n^0$, $\widehat{\boldsymbol{\theta}}_n^k = \text{Arg}\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} W_n(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1})$ for $k = 1,2,\ldots$. Note that, by Lemma 2 of Jennrich (1969), $\widehat{\boldsymbol{\theta}}_n^k$ exists because $W_n(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1})$ is continuous as a function of $\boldsymbol{\theta}$ on the compact space $\boldsymbol{\Theta}$ for k fixed. This process gives rise to a Fisher scoring scheme given by

$$\widehat{\boldsymbol{\theta}}_n^k = \widehat{\boldsymbol{\theta}}_n^{k-1} + [\dot{\Psi}_n^k(\widehat{\boldsymbol{\theta}}_n^{k-1})]^{-1}\Psi_n^k(\widehat{\boldsymbol{\theta}}_n^{k-1}), \quad k = 1,2,3,\ldots, \tag{2.1}$$

where $\Psi_n^k(\boldsymbol{\theta})$ is the rank score function defined by

$$\Psi_n^k(\boldsymbol{\theta}) \equiv W_n'\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{r_i(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1})}{n+1} - \frac{1}{2}\right]\frac{g'(\mathbf{x}_i^t\boldsymbol{\theta})}{\sqrt{\nu(g(\mathbf{x}_i^t\widehat{\boldsymbol{\theta}}_n^{k-1}))}}\mathbf{x}_i \tag{2.2}$$

and $\dot{\Psi}_n^k(\boldsymbol{\theta}) = d\Psi_n^k(\boldsymbol{\theta})/d\boldsymbol{\theta}$. Here $r_i(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1})$ is the rank of $e_i(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1})$ among $e_1(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1}),\ldots,e_n(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1})$, where

$$e_i\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n^{k-1}\right) = \frac{y_i - g(\mathbf{x}_i^t\boldsymbol{\theta})}{\sqrt{\nu(g(\mathbf{x}_i^t\widehat{\boldsymbol{\theta}}_n^{k-1}))}}, \quad i = 1,\ldots,n.$$

The quantity ϕ is removed from the denominator since it has no influence on the ranks and hence the kth step estimator $\widehat{\boldsymbol{\theta}}_n^k$.

As an initial estimator, we propose the *naïve* Wilcoxon estimator

$$\widehat{\boldsymbol{\theta}}_n^0 = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{Arg}\min} W_n^0(\boldsymbol{\theta}),$$

where

$$W_n^0(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R(e_i(\boldsymbol{\theta}))}{n+1} - \frac{1}{2}\right)e_i(\boldsymbol{\theta}) \tag{2.3}$$