Computational Statistics and Data Analysis 54 (2010) 1442–1456

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

An active set algorithm to estimate parameters in generalized linear models with ordered predictors

Kaspar Rufibach *

Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Hirschengraben 84, CH-8001 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 15 February 2009
Received in revised form 14 January 2010
Accepted 14 January 2010
Available online 22 January 2010

Keywords:
Ordered explanatory variable
Constrained estimation
Least-squares
Logistic regression
Cox regression
Active set algorithm
Likelihood ratio test under linear constraints

ABSTRACT

In biomedical studies, researchers are often interested in assessing the association between one or more ordinal explanatory variables and an outcome variable, at the same time adjusting for covariates of any type. The outcome variable may be continuous, binary, or represent censored survival times. In the absence of precise knowledge of the response function, using monotonicity constraints on the ordinal variables improves efficiency in estimating parameters, especially when sample sizes are small. An active set algorithm that can efficiently compute such estimators is proposed, and a characterization of the solution is provided. Having an efficient algorithm at hand is especially relevant when applying likelihood ratio tests in restricted generalized linear models, where one needs the value of the likelihood at the restricted maximizer. The algorithm is illustrated on a real life data set from oncology.

1. Introduction

In many applied problems and especially in biomedical studies, researchers are interested in associating an outcome variable to several explanatory variables, typically via a generalized linear or proportional hazards regression model. Here, the explanatory variables or predictors may be continuous, nominal or ordered. Estimates of regression parameters can be obtained via maximizing a least-squares or (partial) likelihood function. Especially if the number of observations is small to moderate, researchers often encounter noisy estimates of the regression parameters, possibly leading to patterns in the regression estimates that violate the a priori knowledge of a factor being ordered. In order to improve accuracy of estimates and efficiency of overall tests for associations, it is tempting to use the prior knowledge of orderings in some of the regression coefficients.

From a Bayesian perspective, receiving estimators in these type of problems is straightforward using Markov Chain Monte Carlo approaches. Pioneered in a linear model framework by Gelfand et al. (1992), Bayesian approaches have been proposed by Dunson and Herring (2003), Dunson and Neelon (2003) and Robert and Hwang (1996). We also refer to the discussion in the latter two papers. To use Gibbs sampling to get the ordered predictor estimate in logistic regression, Holmes and Held (2006) combine the approach in Gelfand et al. (1992) with an auxiliary variable technique. Note that using e.g. flat priors on the regression coefficient vector $\boldsymbol{\beta}$ it is straightforward to show that the maximum a posteriori estimator is equal to the constrained MLE introduced in Section 2.

Although conceptually straightforward, the implementation of these Bayesian approaches is not without fallacies. For obtaining not only point estimates but also assessing whether parameters are equal or strictly ordered across level of predictors, one needs to borrow from more frequentist approaches and "isotonize" unconstrained parameter estimates (Dunson and Neelon, 2003). Only then one can accommodate "flat regions", i.e. successive estimates for ordered levels that are equal.

* Tel.: +41 0 44 634 46 43; fax: +41 0 44 634 43 86.
E-mail address: kaspar.rufibach@ifspm.uzh.ch.

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.01.014

Although there exists vast literature on frequentist estimation subject to order restrictions (Robertson et al., 1988), estimation in the specific regression model discussed here has gained surprisingly little attention (Mukerjee and Tu, 1995). This may be due to the fact that setting up algorithms in these type of problems is generally difficult (Dunson and Neelon, 2003), and requires approaches that need to be adapted to specific problems, necessitating a vast literature for numerous cases of order restricted estimation. We mention Dykstra and Robertson (1982), Jamshidian (2004), Matthews and Crowther (1998), Tan et al. (2007), Taylor et al. (2007), or Balabdaoui et al. (in press) discussing computation of order restricted estimates in specific regression problems, and Balabdaoui and Wellner (2004), Terlaky and Vial (1998) or Rufibach (2007) for estimation of probability densities under order restrictions. Additionally, generalizations of the pool-adjacent-violaters algorithm (PAVA) to inclusion of continuous isotonic covariates are discussed in Bacchetti (1989), Cheng (2009), Ghosh (2007) and Morton-Jones et al. (2000) in the context of "additive isotonic regression". Estimation in this type of model is usually performed using the cyclical PAVA in connection with backfitting. However, note that we are not in this genuinely semiparametric setting, but rather the number of levels of an ordered factor is given a priori and remains fixed for any number of observations.

Recently, a type of algorithm, which has been around in optimization theory for some decades (Fletcher, 1987), has gained considerable attention in the statistical literature: active set algorithms. Dümbgen et al. (2007) use and generalize such an algorithm to compute a log-concave density not only from i.i.d. but even from censored data. An algorithm similar in spirit is the support reduction algorithm discussed in Groeneboom et al. (2008). The latter authors apply it to the estimation of a convex density and to Gaussian deconvolution. A slight generalization of the support reduction algorithm is used to estimate a convex-shaped hazard function in Jankowski and Wellner (in press). Beran and Dümbgen (in press) extend active set algorithms to the estimation of smooth bimonotone functions. They illustrate their algorithm on regression with two ordered covariates, so also treating the example dealt with in this paper. However, Beran and Dümbgen (in press) only consider least-squares or least absolute deviation estimation, and at most two ordered factors. In this paper, we propose an algorithm for an arbitrary number of ordered factors, and we also provide a characterization of the solution.

A key feature of an active set algorithm is, that although iterative, it terminates after finitely many steps, and that the solution is finally found via an unconstrained optimization. This implicitly implies that, as opposed to some Bayesian approaches (Dunson and Neelon, 2003), the active set algorithm is not hurt if estimates of subsequent levels turn out to be equal. In Section 2 we show that the estimation of a regression function in generalized linear models (GLM) under the above ordered factor restriction can be easily performed using such an active set algorithm.

Optimal scaling. A reviewer drew our attention to optimal scaling, where one seeks to assign numeric values to categorical variables in some optimal way, see e.g. Breiman and Friedman (1985), Gifi (1990) and Hastie and Tibshirani (1990), or applied to modeling interactions in Van Rosmalen et al. (2009). In Gifi (1990, Section 2) categories of the original categorical variables are replaced by "category quantifications", and from then on the variables are considered to be quantitative. Note that in the approach discussed in this paper, one does not necessarily look for an optimal transformation, but rather imposes *a priori knowledge* on a given ordered predictor. In the example analyzed in Section 9 it seems plausible that a higher tumor or nodal stage is associated with a higher risk of experiencing a second primary tumor.

Ordered predictors. While the treatment of quantitative and grouped predictors in regression models is straightforward, we briefly review alternative approaches that can be applied to deal with an ordered explanatory variable z. Let us assume the levels of z are coded as $1, \ldots, k$ where $k \geq 2$ and the levels are increasingly ordered, i.e. $1 \leq \cdots \leq k$.

The most straightforward way to incorporate z as a predictor is simply to ignore the information about the groups and consider it a quantitative variable. This approach implicitly assumes that the group levels represent a true dimension, with intervals measured between adjacent categories that correspond to the chosen coding. If the ordinal values are arbitrarily assigned rather than actually measured, the regression coefficient is then difficult or impossible to interpret.

Supposedly the most prevalent approach to incorporate an ordered predictor z in a regression model is to introduce $k-1$ dummy variables $z_2, \ldots, z_k$ where $z_i = 1\{z = i\}, i = 2, \ldots, k$. This approach ignores the additional knowledge of z having ordered levels, entailing that the estimated parameters $\widehat{\beta}_2, \ldots, \widehat{\beta}_k$ corresponding to the above dummy variables may not be increasingly ordered. This is especially relevant in small sample studies, where noisy estimates may confuse the proper order of dummy variable coefficients.

To simplify interpretation of models, especially when interactions are to be incorporated, researchers sometimes resort to dichotomizing a grouped factor, i.e. introducing only one dummy variable $z_1 = 1\{z \leq l\}$, for some $1 \leq l < k$. Here, the additional knowledge about the ordered levels is not used and may cause a substantial loss of predictive information (Steyerberg, 2009, Section 9.1).

Another choice may be polynomial contrasts. One then introduces new variables $z_i = i^2\{z = i\}, i = 2, \ldots, k$. To avoid correlated estimators β_i and therefore mutually dependent tests when doing variable selection, researchers generally prefer to modify the design matrix in order to get orthogonal polynomial contrasts. The function `as.ordered()` in R (R Development Core Team, 2009) does this by default.

Gertheiss and Tutz (2008) proposed a ridge-regression related approach to perform regression with ordered factors. Consider the predictor z with ordered categories $1, \ldots, k$ and the linear regression model

$$\begin{aligned} \boldsymbol{y} &= \beta_2 \boldsymbol{z}_2 + \cdots + \beta_k \boldsymbol{z}_k + \boldsymbol{\varepsilon} \\ &= \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \end{aligned} \tag{1}$$