Statistical Methodology 24 (2015) 52–68

Contents lists available at ScienceDirect

Statistical Methodology

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

Shrinkage, pretest, and penalty estimators in generalized linear models

Shakhawat Hossain [a,*], S. Ejaz Ahmed [b], Kjell A. Doksum [c]

[a] *Department of Mathematics and Statistics, University of Winnipeg, Winnipeg, MB, Canada*

[b] *Department of Mathematics, Brock University, St. Catharines, ON, Canada*

[c] *Department of Statistics, University of Wisconsin, Madison, WI, USA*

ARTICLE INFO

Article history:
Received 22 August 2013
Received in revised form
2 August 2014
Accepted 13 November 2014
Available online 24 December 2014

Keywords:
Asymptotic risk
Candidate subspaces
Generalized linear models
GLM likelihood ratio test
Linear restrictions
L_1GLM
SCAD
Adaptive L_1GLM
Pretest
Stein type shrinkage

ABSTRACT

We consider estimation in generalized linear models when there are many potential predictors and some of them may not have influence on the response of interest. In the context of two competing models where one model includes all predictors and the other restricts variable coefficients to a candidate linear subspace based on subject matter or prior knowledge, we investigate the relative performances of Stein type shrinkage, pretest, and penalty estimators (L_1GLM, adaptive L_1GLM, and SCAD) with respect to the unrestricted maximum likelihood estimator (MLE). The asymptotic properties of the pretest and shrinkage estimators including the derivation of asymptotic distributional biases and risks are established. In particular, we give conditions under which the shrinkage estimators are asymptotically more efficient than the unrestricted MLE. A Monte Carlo simulation study shows that the mean squared error (MSE) of an adaptive shrinkage estimator is comparable to the MSE of the penalty estimators in many situations and in particular performs better than the penalty estimators when the dimension of the restricted parameter space is large. The Steinian shrinkage and penalty estimators all improve substantially on the unrestricted MLE. A real data set analysis is also presented to compare the suggested methods.

* Corresponding author. Tel.: +1 204 786 9492.
E-mail address: sh.hossain@uwinnipeg.ca (S. Hossain).

http://dx.doi.org/10.1016/j.stamet.2014.11.003

1. Introduction

Generalized linear models (GLMs) are useful tools in the social sciences, biology, medicine, and survival analysis. These models are extensions of linear models that allow for greater flexibility than Gaussian linear models. They are based on an assumed relationship between the mean of the response variable and a linear combination of explanatory variables. Data may be assumed to be from several families of probability distributions, including Bernoulli, normal, binomial, Poisson, negative binomial, and gamma distributions, many of which provide good fits to non-normal error structures.

GLMs model the conditional distribution of a random variable Y given a vector of predictors to follow a distribution in the exponential family using a linear combination $\boldsymbol{x}^\top\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a vector of coefficients. The parameter vector $\boldsymbol{\beta}$ is unknown and we wish to estimate it or to test hypotheses about it. These are usually done by applying the maximum likelihood method and the likelihood ratio test.

In this paper we consider the estimation problem for GLMs when there are many potential predictors and some of them may not have influence on the response of interest. There are possibly extraneous predictors in the model. Thus if the main concern is treatment effect, or the effect of biomarkers, extraneous nuisance variables may be lab effect when several labs are involved, or the age and sex of patients. The analysis will be more precise if "nuisance variables" can be left out of the model. This leads to the consideration of two models: The full model that includes all predictors and possible extraneous variables; and a candidate submodel that includes the predictors of main concern while leaving out extraneous variables. One way to deal with this framework is to use pretest procedures that test whether the coefficients of the extraneous variables are zero and then estimate parameters in the model that include coefficients that are rejected by the test. Another approach is to use Stein type shrinkage estimators where the estimated regression coefficient vector is shrunk in the direction of the candidate subspace. This "soft threshold" modification of the pretest method has been shown to be efficient in various frameworks. Ahmed et al. [1], Hossain et al. [11], Judge, and Judge and Mittelhammaer [14], among others have examined the properties of pretest and shrinkage estimation strategies for parametric and semiparametric linear models.

For a real data example, we use the data analyzed by Park and Hastie [17] (this data set was originally collected by Rossouw et al. [18]). In this example, coronary heart disease may be related to the variables: systolic blood pressure, cumulative tobacco, low density lipoprotein cholesterol, adiposity, family history of heart disease, type-A behavior, obesity, alcohol, and age. Since the prior information is not available here, the shrinkage method uses a two step approach. In the first step, candidate active predictors are selected based on the stepwise or subset selection procedures and the model selection criteria, such as AIC or BIC. It shows that cumulative tobacco, low density lipoprotein cholesterol, family history of heart disease, type-A behavior and age are the most important factors and the effect of the other variables may be ignored.

Returning to the general case, we will consider two models: One where $\boldsymbol{x}^\top\boldsymbol{\beta}$ includes all the x-variables with available data and $\boldsymbol{\beta}$ varies freely in $\Re^k$ and a second model where $\boldsymbol{\beta}$ is in a candidate subspace with $\boldsymbol{\beta}$ restricted to a linear subspace of dimension k_1, $k_1 \le k$. That is, we consider a candidate subspace where an unknown k-dimensional parameter vector $\boldsymbol{\beta}$ satisfies a set of k_2 linear restrictions

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}, \tag{1}$$

where $k_2 = k - k_1$, $\boldsymbol{H}$ is a $k_2 \times k$ matrix of rank $k_2 \le k$, and $\boldsymbol{h}$ is a given $k_2 \times 1$ vector of constants. Since $\boldsymbol{H}$ has rank k_2, the k_2 equations do not contain any redundant information about $\boldsymbol{\beta}$. Next we combine the full and sub-model estimators in an adaptive and asymptotically optimal way. We will consider the special candidate submodel where k_2 of the $\boldsymbol{\beta}$ are zero in Section 4. In this case we can write without loss of generality, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \mathbf{0})^\top$ with $\boldsymbol{\beta}_1$ a k_1 vector. We call this the *variable selection submodel.*

Other procedures that lead to shrinkage and submodels are penalized likelihood methods, such as the least absolute shrinkage and selection operator (LASSO) (Tibshirani [21]), the smoothly clipped absolute deviation method (SCAD) (Fan and Li [8]), and adaptive LASSO (Zou [26]). These methods were proposed for linear models and GLMs. They select a submodel by shrinking some regression