Journal of Multivariate Analysis 151 (2016) 133–150

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Constrained inference in linear regression

Thelge Buddika Peiris [a], Bhaskar Bhattacharya [b,*]

[a] Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609-2280, USA

[b] Department of Mathematics, Southern Illinois University, Carbondale, IL 62901-4408, USA

ARTICLE INFO

Article history:
Received 23 October 2014
Available online 3 August 2016

AMS subject classifications:
62H10
62J05
62F30

Keywords:
Chi-bar square distribution
Confidence interval
Least favorable distribution
Likelihood ratio test
Nonnegative least square

ABSTRACT

Regression analysis is probably one of the most used statistical techniques. We consider the case when the regression function is monotonically changing with some or all of the predictors in a region of interest. Restricted confidence interval for the mean of the regression function is constructed when two predictors are present. Earlier analyses would allow an investigator either to ignore monotonicity altogether or to consider only one predictor at a time but not both simultaneously. The methodologies developed are applied on a real data set to study the effects of patients' age and infection risk on their length of stay in US hospitals.

1. Introduction

1.1. Preliminaries

Consider the standard linear regression model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{Y}$ is an $(n \times 1)$ vector, $\boldsymbol{X}$ is an $(n \times p)$ matrix of rank p, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown parameters, and $\boldsymbol{\epsilon}$ is an $(n \times 1)$ multivariate normal vector of errors with zero mean and covariance matrix $\sigma^2\boldsymbol{I}$. There is a wide range of applications where the sign constraints on regression coefficients are useful. This area of statistical research is known as *non-negative least squares (NNLS)*. In image processing or spectral analysis NNLS is quite well-known, where the signs of the regression parameters can be estimated, or known a priori [2,4,5,7,8,19,23]. NNLS regression can be a useful tool for matrix factorization [10]. The non-negative Garrote [3] uses a sign-constraint, where the signs are derived from an initial estimator as is the positive Lasso [6]. This constraint is particularly relevant when modeling non-negative data, which emerge, e.g., in the form of pixel intensity values of an image, time measurements, histograms or count data, economical quantities such as prices, incomes and growth rates. Non-negativity constraints occur naturally in numerous deconvolution and unmixing problems in diverse fields such as acoustics [14], astronomical imaging [1], genomics [13], proteomics [21], spectroscopy [5] and network tomography [15]; see [4] for a survey.

It is more common in order-restricted regression analysis to consider inference under null hypothesis of the type $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$ versus $\boldsymbol{R}\boldsymbol{\beta} \geq \boldsymbol{r}, \ \boldsymbol{R}\boldsymbol{\beta} \neq \boldsymbol{r}$, for some matrix $\boldsymbol{R}$, vector $\boldsymbol{r}$ [18,20]. Restricted statistical inference in regression analysis under nonnegativity constraints on $\boldsymbol{\beta}$ (NNLS) is rare at best. This emerges when the experimenter believes that the regression

* Corresponding author.
E-mail addresses: tbpeiris@wpi.edu (T.B. Peiris), bhaskar@siu.edu (B. Bhattacharya).

http://dx.doi.org/10.1016/j.jmva.2016.07.008
0047-259X/

function changes monotonically with the predictors (see references above). [16] considered the inference for the mean of the response variable when one predictor variable is present, but their work does not extend to higher dimensional cases in a straightforward manner. In this paper we consider the case of *two* predictors following the same format as theirs. Increasing the number of predictors not only makes practically more useful results but also generates new spaces in null hypothesis parameter region which has no counterpart in lower dimensions (e.g., mixed signs in Section 4). We have used tools from calculus and geometry [9] in our analysis. Graphs are used throughout the paper for illustration, where we have used the convention that arrows on axes indicate to the positive directions. Often we use a (cross-sectional) two-dimensional graph to illustrate a three-dimensional region for clarity or when the three dimensional graph is messy to display.

To obtain confidence intervals we have considered the acceptance regions of corresponding one-sided tests [12,23]. Least favorable distributions are used for calculating the critical values of the tests, however, these distributions are known to be conservative. Restricted likelihood ratio tests (LRT) are used, but it is shown that often these tests perform poorly than a related unrestricted test. In such cases, we have proposed an ad hoc test in similar spirit as in [16] to improve on the LRT.

We have applied our methodology on the SENIC data [11]. The primary objective of the study was to determine whether infection surveillance and control programs have reduced the rates of nosocomial (hospital-acquired) infection in US hospitals. Here we suspect β_1 to be positive and β_2 to be negative because older patients seem to stay longer in hospital and higher infection is associated with shorter hospital stay. Whereas the ordinary regression analysis would ignore this monotonicity information, our analysis implements it. Following [16] one has to consider these important predictors only *one at a time*. Our analysis enables one to consider them *simultaneously*. See Section 7 for data analysis on the example.

1.2. Regression basics

Assuming the first column of $\boldsymbol{X}$ to be all ones, and for two predictor variables X_1, X_2, for a sample of size n, the regression model becomes, $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i,\ 1 \le i \le n$. Let $\hat\beta_0, \hat\beta_1$ and $\hat\beta_2$ be the unrestricted maximum likelihood estimates (MLEs) of β_0, β_1 and β_2 respectively. Let, $S_{x_1}^2 = \sum x_{1i}^2,\ S_{x_2}^2 = \sum x_{2i}^2$ and $S^2 = \sum (Y_i - \hat\beta_0 - \hat\beta_1 x_{1i} - \hat\beta_2 x_{2i})^2/\nu$, where $\nu = n-3$. We assume that the columns of $\boldsymbol{X}$ are orthogonal, that is, $\sum_i x_{1i} = 0,\ \ \sum_i x_{2i} = 0$ and $\sum_i x_{1i}x_{2i} = 0$.

Then it is well known that $\hat\beta_0, \hat\beta_1, \hat\beta_2, S^2$ are mutually independent. Further, $\hat\beta_0 \sim \mathcal{N}(\beta_0,\ \sigma^2/n)$, $\hat\beta_1 \sim \mathcal{N}(\beta_1, \sigma^2/S_{x_1}^2)$, $\hat\beta_2 \sim \mathcal{N}(\beta_2, \sigma^2/S_{x_2}^2)$ and $\nu S^2/\sigma^2 \sim \chi_\nu^2$.

Let $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^\top$ where $\gamma_0 = \sqrt{n}\beta_0,\ \gamma_1 = S_{x_1}\beta_1,\ \gamma_2 = S_{x_2}\beta_2$ then the unrestricted MLE of $\boldsymbol{\gamma}$ is $\hat{\boldsymbol{\gamma}} = (\hat\gamma_0, \hat\gamma_1, \hat\gamma_2)^\top = (\sqrt{n}\hat\beta_0, S_{x_1}\hat\beta_1, S_{x_2}\hat\beta_2)^\top \sim \boldsymbol{\mathcal{N}}_3(\boldsymbol{\gamma}, \sigma^2\boldsymbol{I})$.

Under the constraints $\beta_1 \ge 0,\ \beta_2 \ge 0$, the restricted MLEs of β_i's are given by, $\beta_0^* = \hat\beta_0$, $\beta_1^* = \max\{\hat\beta_1, 0\} = \beta_1^+$, $\beta_2^* = \max\{\hat\beta_2, 0\} = \beta_2^+$. Then the restricted parameter space for $\boldsymbol{\gamma}$ is $\{\boldsymbol{\gamma} : \gamma_0 \in \mathbb{R}, \gamma_1 \ge 0, \gamma_2 \ge 0\}$. The restricted MLEs of $\boldsymbol{\gamma}$ are $\gamma_0^* = \hat\gamma_0$, $\gamma_1^* = \max\{\hat\gamma_1, 0\} = \gamma_1^+$, $\gamma_2^* = \max\{\hat\gamma_2, 0\} = \gamma_2^+$.

The case of σ^2 *known* is considered in Sections 2–5. Section 6 considers σ^2 *unknown* case. We end with some discussion in Section 8. Statistical inference under other combinations of sign restrictions of β_1, β_2 can also be developed similarly. Supplement of this paper contains Lemmas 1–4 with proofs, graphs S1–S3, a chart summarizing the distributions of LRT in limiting cases of (x_{01}, x_{02}), tables of critical values and formulas of confidence intervals in original variables (see [17] for further details). The computer programs needed for the example and calculation of critical values are written in fortran and R (available from the authors on request).

2. Inferences for $\boldsymbol{\beta_0 + \beta_1 x_{01} + \beta_2 x_{02}}$

We consider inferences about the mean function $\mathrm{E}(Y) = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02}$ at predictor variable values (x_{01}, x_{02}) for different possible signs of x_{01} and x_{02}.

2.1. Test for $\beta_0 + \beta_1 x_{01} + \beta_2 x_{02}$ ($x_{01} > 0, x_{02} > 0$)

First we consider the hypotheses,

$$\boldsymbol{G}_0 : \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} \le l, \quad \beta_1 \ge 0,\ \beta_2 \ge 0, \qquad \boldsymbol{G}_1 : \beta_1 \ge 0,\ \beta_2 \ge 0, \tag{2.1}$$

for some $l \in \mathbb{R}$. Using the transformation from $\boldsymbol{\beta}$ to $\boldsymbol{\gamma}$, the constraint $\beta_0 + \beta_1 x_{01} + \beta_2 x_{02} \le l$ in (2.1) becomes, $\frac{\gamma_0}{\sqrt{n}} + \frac{\gamma_1 x_{01}}{S_{x_1}} + \frac{\gamma_2 x_{02}}{S_{x_2}} \le l$, or, $\gamma_2 \le b_1 - c_1\gamma_0 - d_1\gamma_1$, where $b_1 = \frac{l S_{x_2}}{x_{02}}$, $c_1 = \frac{S_{x_2}}{x_{02}\sqrt{n}}$ and $d_1 = \frac{x_{01}S_{x_2}}{x_{02}S_{x_1}}$.

Then, using γ_i hypotheses (2.1) are,

$$\boldsymbol{G}_{01} : 0 \le \gamma_2 \le b_1 - c_1\gamma_0 - d_1\gamma_1, \quad 0 \le \gamma_1, \qquad \boldsymbol{G}_{11} : \gamma_1 \ge 0,\ \gamma_2 \ge 0, \tag{2.2}$$

respectively. To visualize geometrically the sets $\boldsymbol{G}_{01}$ and $\boldsymbol{G}_{11}$ in the $\boldsymbol{\gamma}$ space, let $\boldsymbol{K}$ be the closed convex cone bounded by the hyperplanes $\{c_1\gamma_0 + d_1\gamma_1 + \gamma_2 = 0, \gamma_1 \ge 0, \gamma_2 \ge 0\}$, $\{\gamma_2 = 0, 0 \le \gamma_1 \le \frac{-c_1\gamma_0}{d_1}, \gamma_0 \le 0\}$, and $\{\gamma_1 = 0, 0 \le \gamma_2 \le -c_1\gamma_0, \gamma_0 \le 0\}$ and let $\boldsymbol{L} = (b_1/c_1, 0, 0)$, then $\boldsymbol{G}_{01}$ is the shifted cone $\boldsymbol{K} + \boldsymbol{L}$. Shifting the cone $\boldsymbol{K}$ by b_1/c_1 units along the positive direction