Statistics and Probability Letters 86 (2014) 85–90

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Empirical likelihood test for high dimensional linear models

Liang Peng[a], Yongcheng Qi[b], Ruodu Wang[c,*]

[a] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA

[b] Department of Mathematics and Statistics, University of Minnesota–Duluth, 1117 University Drive, Duluth, MN 55812, USA

[c] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada

ARTICLE INFO

Article history:
Received 13 August 2013
Received in revised form 25 December 2013
Accepted 25 December 2013
Available online 3 January 2014

Keywords:
Empirical likelihood
High-dimensional data
Hypothesis test
Linear model

ABSTRACT

We propose an empirical likelihood method to test whether the coefficients in a possibly high-dimensional linear model are equal to given values. The asymptotic distribution of the test statistic is independent of the number of covariates in the linear model.

1. Introduction

Regression model is a commonly employed technique to model the relationship between responses and covariates. Consider the following classical and also the simplest linear regression model

$$Y_i = \beta^T X_i + \epsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where $\beta = (\beta_1, \dots, \beta_p)^T$ is the vector of unknown parameters, $X_1 = (X_{1,1}, \dots, X_{1,p})^T, \dots, X_n = (X_{n,1}, \dots, X_{n,p})^T$ are independent and identically distributed (iid) random vectors, and $\epsilon_1, \dots, \epsilon_n$ are iid random variables with zero mean and variance σ^2 with ϵ_i's being independent of X_i's. Statistical inference for β can be based on either least squares estimator or M-estimator when p is fixed. When p depends on the sample size n and goes to infinity as $n \to \infty$, Portnoy (1984, 1985) studied the consistency and asymptotic normality of M-estimators for β, which requires that p cannot be too large in comparison with the sample size.

Statistical inference for the linear model (1) is needed for the case when p is of an exponential order of n, motivated by the studies in bioinformatics. To deal with the case when many of β_i's are zero (sparsity), one first selects variables with nonzero $\beta_i' s$ and then makes statistical inference for the selected nonzero $\beta_i' s$. It is not surprising that the order of the number of nonzero $\beta_i' s$ cannot be larger than the optimal one in Portnoy (1985). We refer to Bradic et al. (2011) for more details and references on the ultrahigh dimensional situation. Sparse estimators like the famous Lasso estimator (Tibshirani, 1996) and its extensions (Zou, 2006; Meinshausen, 2007) are very powerful in the setting of sparse alternative. Meinshausen et al. (2009) studied the variable selection for high-dimensional linear regression models.

On the other hand, when the number of nonzero β's is large, new techniques are needed. In contrast to the sparse model and variable selection techniques, we study a general setting in this paper. We consider the problem of testing $H_0 : \beta = \beta_0$ against $H_a : \beta \neq \beta_0$ for a given value $\beta_0 \in \mathbb{R}^p$ when p is either fixed or goes to infinity as $n \to \infty$. In particular, we are

* Corresponding author.
E-mail addresses: peng@math.gatech.edu (L. Peng), yqi@d.umn.edu (Y. Qi), wang@uwaterloo.ca (R. Wang).

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.12.019

interested in the case when the alternative hypothesis has a dense shift (i.e. small shifts in many dimensions instead of large shifts in a few dimensions). When p is fixed, the traditional test is Hotelling's T^2 test, based on the test statistic

$$HT = \frac{1}{\hat{\sigma}^2}(\hat{\beta}-\beta_0)^T\left(\frac{1}{n}\sum_{i=1}^n X_iX_i^T\right)^{-1}(\hat{\beta}-\beta_0), \tag{2}$$

where $\hat{\beta} = (\frac{1}{n}\sum_{i=1}^n X_iX_i^T)^{-1}\frac{1}{n}\sum_{i=1}^n Y_iX_i$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}^TX_i)^2$. It is known that $HT \xrightarrow{d} \chi_p^2$ as $n \to \infty$. However, when p is large, finding the inverse matrix in (2) becomes problematic. To overcome such difficulty, we consider the empirical likelihood method.

As a powerful nonparametric likelihood approach, empirical likelihood test is another useful method. More specifically, define the traditional empirical likelihood function for β as

$$L_n^{(T)}(\beta) = \sup\left\{\prod_{i=1}^n (nq_i) : q_1 \ge 0, \ldots, q_n \ge 0, \sum_{i=1}^n q_i = 1, \sum_{i=1}^n q_i(Y_i - \beta^TX_i)X_i = 0\right\}.$$

Under some regularity conditions, one can show that the Wilks theorem holds, i.e., $-2\log L_n^{(T)}(\beta_0)$ converges in distribution to a chi-square limit with p degrees of freedom. Therefore, the empirical likelihood test can be constructed by using the test statistic $-2\log L_n^{(T)}(\beta)$. See Owen (2001) for more details on empirical likelihood methods. However, the maximization in computing $L_n^{(T)}(\beta)$ becomes nontrivial and even unavailable when p is large; see Chen et al. (2008) for discussions on this phenomenon. Empirical likelihood method for high dimensional data can be found in Chen et al. (2009), Hjort et al. (2009) and Peng and Schick (2013).

In this paper we propose a new empirical likelihood test for testing $H_0 : \beta = \beta_0$ against $H_a : \beta \ne \beta_0$ regardless of fixed or divergent p. We begin with an estimator of $\theta = (\beta_0 - \beta)^T\Sigma^2(\beta_0 - \beta)$ where $\Sigma = \mathbb{E}(X_1X_1^T)$. It is obvious that when Σ is positive definite, testing $H_0 : \beta = \beta_0$ against $H_a : \beta \ne \beta_0$ is equivalent to testing $H_0 : \theta = 0$ against $H_a : \theta \ne 0$. To find such an estimator, we split the data into two parts and introduce an empirical likelihood test based on this estimator. It turns out that the new method works for both fixed and divergent p. The sample splitting method was also used and discussed in Peng et al. (in press) and Wang et al. (2013), where they proposed empirical likelihood tests and jackknife empirical likelihood tests for high-dimensional means. Other methods based on sample splitting techniques for variable selection were discussed in Wasserman and Roeder (2009) and Meinshausen et al. (2009). Note that the purpose of sample splitting in this paper is for testing without variable selection, and hence it is different from their methods.

We organize this paper as follows. Section 2 presents the new methodology and main results. A simulation study is given in Section 3. All proofs are put in Section 4.

2. Methodology

We start by splitting the sample into two groups to get a random sample with mean being $\theta = (\beta_0 - \beta)^T\Sigma^2(\beta_0 - \beta)$, where $\Sigma = \mathbb{E}(X_1X_1^T)$. Put $m = [n/2]$, the integer part of $n/2$, and define $\tilde{X}_i = X_{m+i}$, $\tilde{Y}_i = Y_{i+m}$, $\tilde{\epsilon}_i = \epsilon_{i+m}$,

$$W_i(\beta) = (Y_iX_i - X_iX_i^T\beta)^T(\tilde{Y}_i\tilde{X}_i - \tilde{X}_i\tilde{X}_i^T\beta)$$

for $i = 1, \ldots, m$. Then

$$\mathbb{E}W_i(\beta_0) = \mathbb{E}[(X_iX_i^T(\beta_0 - \beta) + X_i\epsilon_i)^T(\tilde{X}_i\tilde{X}_i^T(\beta_0 - \beta) + \tilde{X}_i\tilde{\epsilon}_i)] = (\beta_0 - \beta)^T\Sigma^2(\beta_0 - \beta).$$

When Σ is positive definite, testing $H_0 : \beta = \beta_0$ against $H_a : \beta \ne \beta_0$ is equivalent to testing $H_0 : \mathbb{E}W_1(\beta_0) = 0$ against $H_a : \mathbb{E}W_1(\beta_0) \ne 0$. This motivates us to apply the empirical likelihood method in Qin and Lawless (1994) to the estimating equation $\mathbb{E}W_1(\beta_0) = 0$. However this direct application results in a poor power in general by noting that $\mathbb{E}W_1(\beta_0) = O(\|\beta - \beta_0\|^2)$ instead of $O(\|\beta - \beta_0\|)$ when $\|\beta - \beta_0\|$ is small, where $\|\cdot\|$ denotes the L_2 norm of a vector. The explanation of this weak power is discussed in Peng et al. (in press).

To improve the power, we propose to add one more linear equation $\mathbb{E}W_1^*(\beta_0) = 0$ where $\mathbb{E}W_1^*(\beta_0)$ is close to $O(\|\beta - \beta_0\|_1)$ where $\|\beta - \beta_0\|_1$ is the L_1 norm, and thus it captures the small change of $\beta - \beta_0$. More specifically, define

$$W_i^*(\beta) = (Y_iX_i - X_iX_i^T\beta)^T\mathbf{1}_p + (\tilde{Y}_i\tilde{X}_i - \tilde{X}_i\tilde{X}_i^T\beta)^T\mathbf{1}_p$$

for $i = 1, \ldots, m$, where $\mathbf{1}_p = (1, 1, \ldots, 1)^T \in \mathbb{R}^p$, and then define the empirical likelihood function for β as

$$L_n(\beta) = \sup\left\{\prod_{i=1}^m (mq_i) : q_1 \ge 0, \ldots, q_m \ge 0, \sum_{i=1}^m q_i = 1, \sum_{i=1}^m q_iW_i(\beta) = 0, \sum_{i=1}^m q_iW_i^*(\beta) = 0\right\}.$$

The following theorem shows that the Wilks theorem holds for the above empirical likelihood method. We use $\mathbf{tr}(A)$ to denote the trace of a matrix A.

Theorem 1. *Let β_0 be the true value of the parameter β. Assume Σ is positive definite and there exists some $\delta > 0$ such that*

$$\frac{\mathbb{E}|X_1^T\tilde{X}_1|^{2+\delta}}{\{\mathbf{tr}(\Sigma^2)\}^{(2+\delta)/2}}\left(\frac{\mathbb{E}|\epsilon_1|^{2+\delta}}{\sigma^{2+\delta}}\right)^2 = o(m^{\delta/2}), \tag{3}$$