Applied Mathematics and Computation 000 (2017) 1–9

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Optimal estimation of direction in regression models with large number of parameters

Jonathan Gillard[a,*], Anatoly Zhigljavsky[a,b]

[a] School of Mathematics, Cardiff University, Cardiff CF24 4AG, UK
[b] Lobachevsky Nizhny Novgorod State University, Nizhny Novgorod 603950, Russia

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Random balance
Screening experiments
Box–Wilson methodology
LASSO
Ridge regression

ABSTRACT

We consider the problem of estimating the optimal direction in regression by maximizing the probability that the scalar product between the vector of unknown parameters and the chosen direction is positive. The estimator maximizing this probability is simple in form, and is especially useful for situations where the number of parameters is much larger than the number of observations. We provide examples which show that this estimator is superior to state-of-the-art methods such as the LASSO for estimating the optimal direction.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

In this paper, we are mainly interested in the problem of choosing the optimal direction in regression by maximizing the probability that the scalar product between the vector of unknown parameters and the chosen direction is positive. The results obtained are very general and could be applied to models where the number of parameters m exceeds the number of observations N. It turns out that the optimal directional vector has a very simple form, see (3), and can be easily computed even if the number of parameters m is extremely large. There are two very important practical areas where our directional statistic, denoted $\hat{\theta}_*$, can be used; see also Sections 5.1 and 5.2.

- The Box–Wilson response surface methodology, see [1,2] and [3, Ch. 8A], where an unknown response function can be observed with random error and the aim of the experimentation is in reaching the experimental conditions where the response function achieves its maximum. The main step (applied many times) in this methodology is the construction of a local linear model of the response function and the estimation of the coefficients of this linear model for finding the direction of ascent. The standard advice is to use the LSE (least square estimator) for estimating the coefficients. As shown in this paper, this standard procedure can be much improved as the LSE does not provide the optimal direction. Also, the use of $\hat{\theta}_*$ in place of the LSE can expand the use of the Box–Wilson methodology to problems with very large number of input variables.
- The so-called 'sure independence screening' procedure for regression models with huge number of parameters, see [4] as a classical reference. This procedure consists of two stages. At the first stage, a computationally efficient method is used for screening out the most important variables quickly, thus reducing the dimensionality. At the second stage, a proper regression analysis is applied to the remaining variables. Our arguments show that $\hat{\theta}_*$ is not only computationally simple

[*] Corresponding author.
E-mail addresses: gillardjw@cardiff.ac.uk, gillardjw@cf.ac.uk (J. Gillard), zhigljavskyaa@cardiff.ac.uk (A. Zhigljavsky).

http://dx.doi.org/10.1016/j.amc.2017.05.050
0096-3003/© 2017 Elsevier Inc. All rights reserved.

but also provides an optimal screening procedure to be applied at the first stage of the sure independence screening approach.

Assume we have N observations in the linear regression model

$$y_j = \theta_1 x_{j1} + \cdots + \theta_m x_{jm} + \varepsilon_j, \quad j = 1, \ldots, N. \tag{1}$$

In a standard way (see e.g. [5, Ch. 4]), we write the matrix version of this observation scheme as

$$Y = X\theta + \varepsilon \tag{2}$$

where $Y = (y_1, \ldots, y_N)^T$ is the observation vector (response variable), $X = (x_{ji})_{j,i=1}^{N,m}$ is the design matrix, $\theta = (\theta_1, \ldots, \theta_m)^T$ is the vector of unknown parameters and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)^T$ is a vector of noise. As usual in regression models we assume $\mathbb{E}\varepsilon = 0$ and the covariance matrix of errors is $D\varepsilon = \sigma^2 W$, where σ^2 is generally unknown and W is some positive definite $N \times N$ matrix. In Section 2 we assume that W is the identity $N \times N$ matrix (that is, $W = I_N$) and extend the main results to the general $W > 0$ in Section 2.2.

The main result of the paper is Theorem 2.1 which states that if $Y \sim N(0, \sigma^2 I_N)$ then the statistic

$$\hat{\theta}_* = X^T Y \tag{3}$$

maximizes the probability

$$\Pr\{v^T \theta_{\text{true}} > 0\} \tag{4}$$

over all vectors $v \in \mathbb{R}^m$, where θ_{true} is the true value of the unknown parameters θ.

Let us make two important remarks.

Remark 1. For any vector v, the probability (4) is the same for all vectors γv with $\gamma > 0$. This means that our focus is solely on the directions generated by vectors $v \in \mathbb{R}^m$ rather than on the estimation of $\theta = \theta_{\text{true}}$ in the regression model (2). Moreover, Theorem 2.1 implies that under appropriate assumptions all estimators of the form $\gamma X^T Y$ with $\gamma > 0$ are optimal with respect to the criterion (4).

Remark 2. Careful examination of the proof of Theorem 2.1 shows that for given $\theta = \theta_{\text{true}}$ there could be other directions optimal for the criterion (4). A remarkable property of the direction defined by $\hat{\theta}_*$ is the fact that this direction is optimal for any θ_{true}. We can state this property by saying that $\hat{\theta}_*$ is universally optimal with respect to the criterion (4).

The rest of the paper is organized as follows. In Section 2 we prove our main result, Theorem 2.1, and show how this result can be further generalized and used. In Section 3 we give two analytic examples which show that the direction created by $\hat{\theta}_*$ could be much superior to the direction generated by the BLUE and other linear estimators of θ. In Section 4 we provide results of several numerical studies which further confirm the superiority of $\hat{\theta}_*$. As a by-product of the numerical study of Section 4 we show that the celebrated LASSO can perform very poorly in terms of the criterion (4). We make further discussions in Section 5, where we also formulate conclusions.

2. Optimality of the directional statistic

2.1. The main result

In a general linear regression model (2), consider a family of linear statistics of the form

$$\hat{\theta}_C = CY\,, \tag{5}$$

where C is some $m \times N$ matrix. Define the scalar product in $\mathbb{R}^m$ by

$$\langle a, b \rangle = a^T S b\,, \quad a, b \in \mathbb{R}^m\,,$$

where S is an arbitrary positive definite $m \times m$ matrix. For given θ, define

$$\mathcal{C}_\theta = \text{Argmax}_C \Pr\{\langle \hat{\theta}_C, \theta \rangle > 0\}; \tag{6}$$

that is, $\mathcal{C}_\theta = \{C_\star\}$ is the set of $m \times N$ matrices $C_\star$ such that

$$\Pr\{\langle \hat{\theta}_{C_\star}, \theta \rangle > 0\} = \max_C \Pr\{\langle \hat{\theta}_C, \theta \rangle > 0\}\,. \tag{7}$$

For given θ, we say that a statistic $\hat{\theta}_C$ is optimal if $C \in \mathcal{C}_\theta$. The theorem below shows that if we assume normality of errors then the matrix $C_* = S^{-1} X^T \in \mathcal{C}_\theta$ for all θ. This matrix does not depend on θ and, if $S = I_m$, the corresponding optimal statistic $\hat{\theta}_{C_\star}$ coincides with $\hat{\theta}_\star$ defined in (3).

Theorem 2.1. *Consider the model (2) where $\varepsilon \sim N(0, \sigma^2 I_N)$, $\sigma^2 > 0$, and let S be any positive definite $m \times m$ matrix. Then for any θ, $C_\star = S^{-1} X^T$ belongs to the set $\mathcal{C}_\theta$ defined in (6).*

Please cite this article as: J. Gillard, A. Zhigljavsky, Optimal estimation of direction in regression models with large number of parameters, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.05.050

Download English Version:

https://daneshyari.com/en/article/8901596

Download Persian Version:

https://daneshyari.com/article/8901596

Daneshyari.com