Statistics and Probability Letters 93 (2014) 1–6

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Shrinkage estimator in normal mean vector estimation based on conditional maximum likelihood estimators

Junyong Park

Department of Mathematics and Statistics, University of Maryland Baltimore County, MD 21250, United States

ARTICLE INFO

Article history:
Received 4 October 2013
Received in revised form 1 June 2014
Accepted 1 June 2014
Available online 12 June 2014

Keywords:
Shrinkage
Sparsity
Empirical Bayes
Conditional maximum likelihood estimate
Mean vector
Stein's risk unbiased estimate

ABSTRACT

Estimation of normal mean vector has broad applications such as small area estimation, estimation of nonparametric functions and estimation of wavelet coefficients. In this paper, we propose a new shrinkage estimator based on conditional maximum likelihood estimator incorporating with Stein's risk unbiased estimator (SURE) when data have the normality. We present some theoretical work and provide numerical studies to compare with some existing methods.

1. Introduction

Let $\mathbf{z} = (z_1, \ldots, z_n)$ be a vector of observations generated independently from $z_i = \theta_i + \epsilon_i,\ i = 1, 2, \ldots, n$ where $\epsilon_i \sim N(0, \sigma^2)$. Our interest is to obtain an estimate of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, say $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$, with small L_2 risk $R(\boldsymbol{\theta}) = \sum_{i=1}^{k} E(\hat{\theta}_i - \theta_i)^2$. In particular, when the solutions are sparse in that the number of nonzero θ's is small compared to n, a variety of shrinkage estimators have been developed, for example, see Donoho and Johnstone (1994), Donoho and Johnstone (1995), and Johnstone and Silverman (2004). The sparsity situation occurs commonly in many practical problems, for example, the estimation of wavelet coefficients which is also known as wavelet denoising. Under the sparsity, these literatures discussed the use of different types of shrinkage such as hard shrinkage and soft shrinkage. These two shrinkage estimators are the most typical examples when only a small proportion of θ_i is non-zero. The hard and soft shrinkage have some drawbacks in practice, for example, hard shrinkage has discontinuity in estimator causing a large variability and the soft shrinkage has large bias. Johnstone and Silverman (2004) proposed the empirical Bayes (EB) approach with different prior distributions on θ which has the form of mixture of point mass at 0 and the nonzero part.

In this paper, we propose an adaptive and model-based shrinkage estimator of which the form reflect the normality. This will be implemented based on conditional maximum likelihood estimate (CMLE) conditioning on observations greater or smaller than some cut-off value. Our proposed CMLE is also expected to overcome drawbacks of hard and soft thresholding such as discontinuity of hard thresholding and large bias of soft thresholding. We incorporate the idea of SURE (Stein's Unbiased Risk Estimate) to determine the parameter in the proposed shrinkage estimator as in the case of soft shrinkage estimator with SURE. However it will be seen that our proposed CMLE with SURE performs better than the soft shrinkage with SURE. Our proposed CMLE is a fully data-dependent procedure whereas the EB approach in Johnstone and Silverman (2004) needs to specify the prior distribution. We also provide some optimal property of the proposed CMLE as Donoho and

E-mail address: junpark@umbc.edu.

http://dx.doi.org/10.1016/j.spl.2014.06.005
0167-7152/

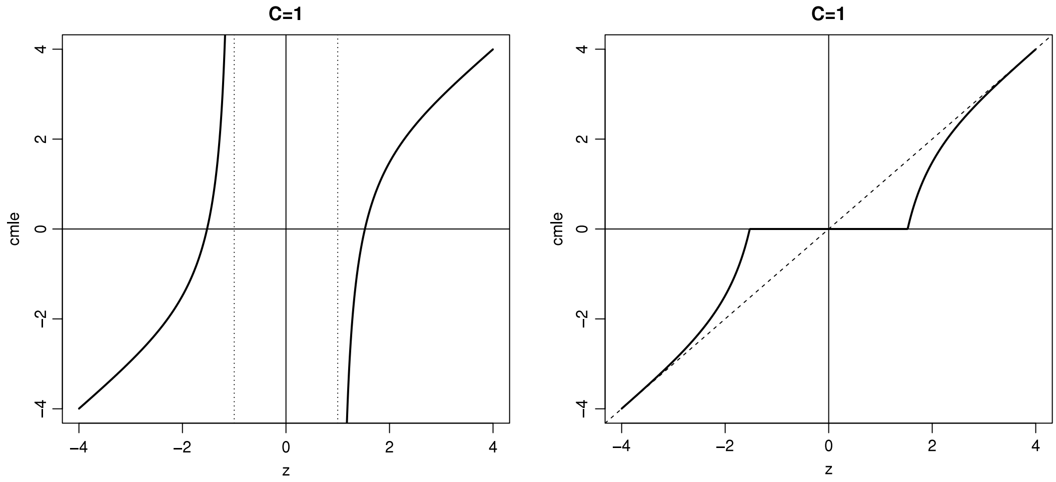

Fig. 1. The left panel in the first row shows $\hat{\theta}^{\text{CMLE}}$ conditioning on $Z_j > 1$ or $Z_j < -1$ and the right panel shows $\hat{\theta}$ by modifying $\hat{\theta}^{\text{CMLE}}$.

Johnstone (1994) did for hard and soft shrinkage and present numerical studies showing that our proposed CMLE obtains at least competitive or better performances than EB as well as hard and soft shrinkages.

This paper is organized as follows. In Section 2, we introduce our proposed CMLE and present some property to reduce computation. Section 3 provides some discussion on the Stein's risk unbiased estimate to determine the shrinkage parameter. In Section 4, numerical studies are presented to compare the CMLE and some other methods. We present concluding remarks in Section 5.

2. Conditional maximum likelihood estimate

In this section, we propose an estimator of $\boldsymbol{\theta}$ based on conditional MLE. The conditional densities of Z_j given $Z_j > C$ and $Z_j < -C$ are

$$\frac{\frac{1}{\sigma}\phi\left(\frac{z_j-\theta_j}{\sigma}\right)}{1-\Phi\left(\frac{C-\theta_j}{\sigma}\right)} I(z_j > C), \qquad \frac{\frac{1}{\sigma}\phi\left(\frac{z_j-\theta_j}{\sigma}\right)}{\Phi\left(\frac{-C-\theta_j}{\sigma}\right)} I(z_j < -C). \tag{1}$$

We consider the CMLE conditioning on either $Z_j > C$ or $Z_j < -C$, namely $\hat{\theta}_j^{\text{CMLE}}$ which is

$$\hat{\theta}_j^{\text{CMLE}} = \begin{cases} \text{argmax}_{\theta_j} \dfrac{\frac{1}{\sigma}\phi\left((z_j-\theta_j)/\sigma\right)}{1-\Phi\left((C-\theta_j)/\sigma\right)} & \text{if } z_j > C \\ 0 & \text{if } |z_j| \le C \\ \text{argmax}_{\theta_j} \dfrac{\frac{1}{\sigma}\phi\left((z_j-\theta_j)/\sigma\right)}{\Phi\left((-C-\theta_j)/\sigma\right)} & \text{if } z_j < -C. \end{cases}$$

However, as shown in the left graphs in Fig. 1 for $C = 1$, this $\hat{\theta}^{\text{CMLE}}$ has negative (positive) estimators even when $Z_j > 0$ ($Z_j < 0$), respectively. In order to avoid this property and to obtain continuity, we modify the $\hat{\theta}^{\text{CMLE}}$ further to have our proposed estimator, namely $\hat{\theta}$, as follows:

$$\hat{\theta}_j \equiv \hat{\theta}(Z_j) = \begin{cases} \max(\hat{\theta}_j^{\text{CMLE}}, 0) & \text{if } Z_j > 0 \\ \min(\hat{\theta}_j^{\text{CMLE}}, 0) & \text{otherwise.} \end{cases}$$

As shown from two right graphs in Fig. 1, $\hat{\theta}$ from $\hat{\theta}^{\text{CMLE}}$ is smooth (in the sense of weakly differentiable).

We see that $\hat{\theta}_j$'s have the sparse solution in that we have 0 estimators for relatively small observation Z_js. Furthermore, as mentioned, the estimator has the continuity so we can apply the idea of SURE to determine the parameter C. We demonstrate this estimation of C in the next section.

By symmetry, we only need to consider the case of $Z_j > C$ since the other case can be done easily. Given an observation $Z_j > C$ and σ, the $\hat{\theta}_j^{\text{CMLE}}$ is the solution of

$$Z_j = \theta_j + \sigma \cdot h\left(\frac{C-\theta_j}{\sigma}\right) \tag{2}$$