Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 684–696
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Adaptation over parametric families of symmetric linear estimators☆

Rudolf Beran*

Department of Statistics, University of California, Davis, CA 95616, USA

Available online 28 July 2006

Abstract

This paper treats an abstract parametric family of symmetric linear estimators for the mean vector of a standard linear model. The estimator in this family that has smallest *estimated* quadratic risk is shown to attain, asymptotically, the smallest risk achievable over all candidate estimators in the family. The asymptotic analysis is carried out under a strong Gauss–Markov form of the linear model in which the dimension of the regression space tends to infinity. Leading examples to which the results apply include: (a) penalized least squares fits constrained by multiple, weighted, quadratic penalties; and (b) running, symmetrically weighted, means. In both instances, the weights define a parameter vector whose natural domain is a continuum.

MSC: Primary 62G08; secondary 62G20

Keywords: Biased estimator; Estimated risk; Dimensional asymptotics; Penalized least squares; Running weighted average; Regularization; Linear model

1. Introduction

Consider the standard linear model in which the $n \times 1$ observation vector y satisfies

$$y = \eta + e, \quad \eta = X\beta. \tag{1.1}$$

Here, the design matrix X is $n \times p$ and has rank p, the $p \times 1$ vector β is unknown, and the components of the $n \times 1$ vector e are independent, identically distributed with mean zero, unknown variance σ^2, and finite fourth moment. For brevity, we will call this the *strong Gauss–Markov model*. The problem of estimating η well under this model has motivated major developments in statistical theory and practice.

For any matrix A, including the special case of a vector, let $|A|$ denote the Euclidean (or Frobenius) norm: $|A|^2 = \mathrm{tr}(A'A) = \mathrm{tr}(AA')$. Define the normalized quadratic loss of any estimator $\hat{\eta}$ of η to be

$$L(\hat{\eta}, \eta) = p^{-1}|\hat{\eta} - \eta|^2. \tag{1.2}$$

☆ This research was supported in part by National Science Foundation Grant DMS 0404547.
* Tel.: +1 530 754 7765; fax: +1 530 750 7099.
E-mail address: beran@wald.ucdavis.edu.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.06.029

The risk of $\hat{\eta}$ is then

$$R(\hat{\eta}, \eta, \sigma^2) = EL(\hat{\eta}, \eta), \tag{1.3}$$

where the expectation is calculated under the strong Gauss–Markov model (1.1).

A *linear estimator* of η has the form $\hat{\eta} = Ay$, where A is a $n \times n$ matrix that does not depend on y. It may be biased or unbiased for η. The least squares estimator, $\hat{\eta}_{\mathrm{LS}} = X(X'X)^{-1}X'y$, is an unbiased linear estimator with risk $R(\hat{\eta}_{\mathrm{LS}}, \eta, \sigma^2) = \sigma^2$. According to the Gauss–Markov theorem, $\hat{\eta}_{\mathrm{LS}}$ has smallest risk among all linear unbiased estimators of η. However, Stein (1956) proved that the least squares estimator is inadmissible for η under quadratic loss whenever $p \geqslant 3$ and the errors are independent, identically normally distributed. In statistical practice, $\hat{\eta}_{\mathrm{LS}}$ is often too variable an estimator unless the number p of regressors is small.

These findings have led to consideration of biased linear estimators for η. The risk function of the linear estimator $\hat{\eta} = Ay$ is

$$R(Ay, \eta, \sigma^2) = p^{-1}[\sigma^2 \operatorname{tr}(A'A) + \eta'(I_n - A)'(I_n - A)\eta]. \tag{1.4}$$

This risk is a convex function of A. Its form suggests the possibility of reducing risk through trade-off, by choice of A, between the variance terms $\sigma^2 \operatorname{tr}(A'A)$ and the bias term $\eta'(I_n - A)'(I_n - A)\eta$.

We find the matrix $\tilde{A}$ that minimizes the risk (1.4). Because

$$R(Ay, \eta, \sigma^2) = p^{-1}[\sigma^2 \operatorname{tr}(A'A) + \eta'\eta - 2\eta' A\eta + \eta' A' A\eta] \tag{1.5}$$

and the matrix derivatives

$$\partial \operatorname{tr}(A'A)/\partial A = 2A, \quad \partial\eta' A\eta/\partial A = \eta\eta', \quad \partial\eta' A' A\eta/\partial A = 2A\eta\eta' \tag{1.6}$$

(cf. Section A.15 in Rao and Toutenberg, 1995), it follows that:

$$\partial R(Ay, \eta, \sigma^2)/\partial A = p^{-1}[2\sigma^2 A - 2\eta\eta' + 2A\eta\eta']. \tag{1.7}$$

Setting this risk derivative equal to zero and simplifying yields

$$\tilde{A} = I_n - (I_n + \sigma^{-2}\eta\eta')^{-1} = (\sigma^2 + |\eta|^2)^{-1}\eta\eta'. \tag{1.8}$$

Let

$$H = X'X, \quad U = XH^{-1/2}. \tag{1.9}$$

Evidently U is $n \times p$ and $U'U = I_p$. It follows from (1.9) that $\eta = U\xi$ with $\xi = H^{1/2}\beta$. Consequently,

$$\tilde{A} = U\tilde{S}U', \quad \tilde{S} = (\sigma^2 + |\xi|^2)^{-1}\xi\xi'. \tag{1.10}$$

Note that $\tilde{S}$ is $p \times p$ symmetric with all eigenvalues between 0 and 1. The oracle linear estimator $\tilde{A}y$ minimizes risk among all linear estimators Ay. It is usually not realizable because we usually lack accurate knowledge of $\eta\eta'$ and σ^2. However, representation (1.10) indicates that, among linear estimators of η, we may reasonably restrict attention to those having the form USU', where S is a $p \times p$ symmetric matrix with all eigenvalues between 0 and 1 and U is given by (1.9). The least squares estimator $\hat{\eta}_{\mathrm{LS}}$ has this form with $S = I_p$.

An extensive literature, reviewed in Buja et al. (1989) and in Kneip (1994), has developed specific examples of symmetric linear estimators that can reduce risk by smoothing or shrinkage, with submodel fitting as a limiting case. In considering this work, it is important for understanding to distinguish between two problems: estimation of the discrete vector η versus estimation of a function that coincides with η at certain design points. The first problem, estimating discrete η, arises in analyzing discrete complete or incomplete multi-way layouts, including regression. Solutions to the discrete problem do not require existence or smoothness of an interpolating function that is to be estimated.