Minimax risk over $l_p$-balls for $l_q$-error

David L. Donoho, Iain M. Johnstone
Department of Statistics, Stanford University, Stanford, CA 94305, USA

Received: 23 June 1992/In revised form: 13 November 1993

Summary. Consider estimating the mean vector $\theta$ from data $N_n(\theta, \sigma^2 I)$ with $l_q$ norm loss, $q \geq 1$, when $\theta$ is known to lie in an $n$-dimensional $l_p$ ball, $p \in (0, \infty)$. For large $n$, the ratio of minimax linear risk to minimax risk can be arbitrarily large if $p < q$. Obvious exceptions aside, the limiting ratio equals 1 only if $p = q = 2$. Our arguments are mostly indirect, involving a reduction to a univariate Bayes minimax problem. When $p < q$, simple non-linear co-ordinatewise threshold rules are asymptotically minimax at small signal-to-noise ratios, and within a bounded factor of asymptotic minimaxity in general. We also give asymptotic evaluations of the minimax linear risk. Our results are basic to a theory of estimation in Besov spaces using wavelet bases (to appear elsewhere).

Mathematics Subject Classification (1985): 62C20, 62F12, 62G20

1 Introduction

Suppose we observe $y = (y_i)_{i=1}^n$ with $y_i = \theta_i + z_i$, $z_i$ i.i.d. $N(0, \sigma^2)$, with $\theta = (\theta_i)_{i=1}^n$ an unknown element of the convex set $\Theta$. Sacks and Strawderman (1982) showed that, in some cases, the minimax linear estimator of a linear functional $L(\theta)$ could be improved on by a nonlinear estimator. Specifically, they showed that for squared error loss, the ratio $R_\Theta^L/R_\Theta^N$ of minimax risk among linear estimates to minimax risk among all estimates exceeded $1 + \varepsilon$ for some (unknown) $\varepsilon > 0$ depending on the problem. This raised the possibility that nonlinear estimators could dramatically improve on linear estimators in some cases.

However, Ibragimov and Hasminskii (1984) established a certain limitation on this possibility by showing that there is a positive finite constant bounding the ratio $R_\Theta^L/R_\Theta^N$ for any problem where $\Theta$ is symmetric and convex. Donoho, et al. (1990) have shown that the Ibragimov–Hasminskii constant is not larger than 5/4. Moreover, Donoho and Liu (1991) have shown that even if $\Theta$ is convex but asymmetric, still $R_\Theta^L/R_\Theta^N < 5/4$—provided inhomogeneous linear estimators are allowed. It follows that for estimating a single linear functional, minimax linear estimates cannot be dramatically improved on in the worst case.
Some results for $\ell_2$ error

For the problem of estimating the whole object $\theta$, with squared $l_2$-loss $\|\hat{\theta} - \theta\|^2 = \sum (\hat{\theta}_i - \theta_i)^2$, one could ask again whether linear estimates are nearly minimax. Pinsker (1980) discovered that if $\Theta$ is an ellipsoid, then $R_L^*/R_N^* \to 1$ as $n \to \infty$. Donoho, et al. (1990) showed that if $\Theta$ is an $l_p$-body with $p \geq 2$ then $R_L^*/R_N^* \leq 5/4$, nonasymptotically. Thus there are again certain limits on the extent to which nonlinear estimates can improve on linear ones in the worst case.

However, these limits are less universal in the case of estimating the whole object than they are in the case of estimating a single linear functional. In this paper we show that there are cases where the ratio $R_L^*/R_N^*$ may be arbitrarily large. We begin by highlighting some conclusions for the case of $\ell_2$-error, and give later a systematic description of more general results for $\ell_p$-error. Let $\Theta_{p,n}$ denote the standard $n$-dimensional unit ball of $l_p$, i.e. $\Theta_{p,n} = \{ \theta : \sum |\theta_i|^p \leq 1 \}$.

**Theorem 1** Let $n \sigma^2(n) = \text{constant}$ and $\Theta = \Theta_{p,n}$. Then as $n \to \infty$

$$\frac{R_L^*}{R_N^*} \to \begin{cases} 1 & p \geq 2 \\ \infty & p < 2 \end{cases} \tag{1}$$

This reflects the phenomenon that in some function estimation problems of a linear nature, the optimal rate of convergence over certain convex function classes is not attained by any linear estimate (Kernel, Spline, ...). Compare also Sects. 7–9 in Donoho, et al. (1990), and the discussion below.

Our technique sheds some light on this phenomenon of Pinsker's. It shows

**Theorem 2** Let $p$ be fixed, and set $\Theta = \Theta_{p,n}$. Suppose that we can choose $\sigma^2 = \sigma^2(n)$ in such a way that $R_L^*/R_N^* \to 1$. There are 3 possibilities:

a. $R_L^*/n \sigma^2 \to 1$ (Classical case).

b. $p = 2$ (Pinsker's case).

c. $R_L^*/n \sigma^2 \to 0$ (trivial case).

In words, if the minimax linear estimator is nearly minimax, then: either (case a) the raw data $y$ is nearly minimax, or (case c) the trivial estimator $\theta$ is nearly minimax, or else we are in the case $p = 2$ covered by Pinsker (1980). Put differently, Pinsker's phenomenon happens among $l_p$ constraints only if $p = 2$.

Theorems 1 and 2 show that improvement on minimax linear estimation is possible without showing how (or by how much). A heuristic argument suggests that a non-linear estimator that is near optimal has the form

$$\hat{\theta}_{\lambda, t} = \text{sgn}(y_i)(|y_i| - \lambda \sigma)_+ \tag{2}$$

where $\lambda = \lambda(n, \sigma, p)$. Consider, for example, the case $p = 1$ and $\sigma = cn^{-1/2}$. Then, on average $|\theta_i| \leq n^{-1}$. Therefore most of the coordinates $\theta_i$ are of order $n^{-1}$ in magnitude. But for Theorem 1, $\sigma = O(n^{-1/2})$. The nonlinear estimator with $\lambda = 5 \cdot \sigma$ will estimate most coordinates as 0 and so will be wrong in most coordinates by only $O_p(n^{-1})$, and the case of the few others by only $O_p(n^{-1/2})$. As the minimax linear estimator is wrong in every coordinate by $O_p(n^{-1/2})$, the result (1) for $p < 2$ might not be surprising.

In fact, for an appropriate choice of $\lambda$, the estimator (2) is asymptotically minimax, and the improvement $R_L^*/R_N^*$ can be calculated.