L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

On computation of regression quantiles: Making the Laplacian Tortoise faster

Stephen Portnoy

Department of Statistics, University of Illinois, Champaign, USA

Abstract: In "The Gaussian Hare and the Laplacian Tortoise", the authors present a two-pronged attacked on the computation of L_1 and other regression quantile estimators in linear models for large samples. The first prong involves the application of interior point linear programming methods, specifically designed to treat the absolute error and related regression quantile objective functions. The second prong applies a form of stochastic preprocessing, somewhat reminiscent of the O(n) algorithms for computing the median of a single sample. These ideas provide computational methods that are in theory faster than least squares as $n \to \infty$ (with probability tending to one), and in practice are faster than Splus least squares functions for n larger than 10^4 (and the number of parameters moderate). Here some issues concerning this algorithm are considered, and some improvements are proffered.

Key words: Linear models, regression quantiles, L_1-estimation, computation.

AMS subject classification: 62G05, 62J05, 68A20.

1 Introduction

Consider the standard linear model: there are observations $\{Y_i : i = 1, ..., n\}$, satisfying

$$Y_i = x_i'\beta + u_i \qquad i = 1, ..., n \tag{1}$$

where x_i are vectors in $\mathbf{R}^p$, $\beta \in \mathbf{R}^p$ is a vector of unknown parameters, and $\{u_i\}$ form an i.i.d. sequence of errors. Though we consider the model conditionally on $\{x_i\}$, we will generally assume that these design vectors are realizations of an independent random process. The traditional approach to