A Finite Smoothing Algorithm for Quantile Regression

Colin Chen

[1] R4117 Application, SAS Institute Inc., Cary, NC 27513

Keywords: Quantile Regression, Heteroscedasticity, Finite Smoothing Algorithm, Simplex, Interior Point, Median Regression, Linear Programming

1 Introduction

Quantile regression, which was introduced by Koenker and Bassett (1978), estimates and conducts inference about conditional quantile functions.

A real valued random variable may be characterized by its distribution function,

$$F(y) = \text{ Prob } (Y \leq y)$$

while for any $0 < \tau < 1$,

$$Q(\tau) = \text{ inf } \{y : F(y) \geq \tau\}$$

is called the τ-th quantile of Y. The median, $Q(1/2)$, plays the central role. Like the distribution function, the quantile function provides a complete characterization of the random variable, Y.

For a random sample $\{y_1, ..., y_n\}$ of Y, it is well known that the sample median is the solution of the optimization problem

$$\min_{\xi \in \boldsymbol{R}} \sum_{i=1}^{n} |y_i - \xi|.$$

The general τ-th sample quantile $\xi(\tau)$, which is an analogue of $Q(\tau)$, may be formulated as the solution of the optimization problem

$$\min_{\xi \in \boldsymbol{R}} \sum_{i=1}^{n} \rho_\tau(y_i - \xi),$$

where $\rho_\tau(z) = z(\tau - I(z < 0))$, $0 < \tau < 1$.

As estimating the unconditional mean, viewed as the minimizer,

$$\hat{\mu} = \text{argmin}_{\mu \in \boldsymbol{R}} \sum (y_i - \mu)^2$$

can be extended to estimation of the linear conditional mean function $E(Y|X = x) = x^{'}\beta$ by solving

$$\hat{\beta} = \text{argmin}_{\beta \in \boldsymbol{R}^p} \sum (y_i - x_i^{'}\beta)^2,$$

the linear conditional quantile function, $Q_\tau(\tau|X = x) = x_i^{'}\beta(\tau)$, can be estimated by solving

$$\hat{\beta}(\tau) = \text{argmin}_{\beta \in \boldsymbol{R}^p} \sum \rho_\tau(y_i - x_i^{'}\beta).$$

The median case, $\tau = 1/2$, which is equivalent to minimizing the sum of absolute values, is usually known as the L_1 regression.

An LP Problem

Let $\mu = [y - X\beta]_+$, $\nu = [X\beta - y]_+$, $\phi = [\beta]_+$, and $\varphi = [-\beta]_+$, where $[z]_+$ is the nonegative part of z.

Let $D_{LAR}(\beta) = \sum_{i=1}^{n} |y_i - x_i^{'}\beta|$ and $D_{\rho_\tau}(\beta) = \sum_{i=1}^{n} \rho_\tau(y_i - x_i^{'}\beta)$. For the L_1 problem, the simplex approach solves $\min_\theta D_{LAR}(\theta)$ by the reformulation

$$\min_{\beta}\{e^{'}\mu + e^{'}\nu | y = X\beta + \mu - \nu, \{\mu, \nu\} \in \boldsymbol{R}_{+}^{n}\},$$

where e denotes an n-vector of ones.

Let $B = [X \;\; -X \;\; I \;\; -I]$, $\theta = (\phi \;\; \varphi \;\; \mu \;\; \nu)^{'}$, and $d = (\mathbf{0}^{'} \;\; \mathbf{0}^{'} \;\; e^{'} \;\; e^{'})$ where $\mathbf{0}^{'} = (0\ 0\ \ldots\ 0)_p$. The reformulation presents a standard LP problem:

$$\begin{aligned} &\min_{\theta} d^{'}\theta \\ &B\theta = y \\ &\theta \geq 0. \end{aligned}$$

This problem has the dual formulation

$$\begin{aligned} &\max_{d} y^{'}z \\ &B^{'}z \leq d. \end{aligned}$$

It can be simplified as

$$\max_{z}\{y^{'}z | X^{'}z = 0, z \in [-1, 1]^n\},$$

or, equivalently,

$$\min_{z}\{y^{'}z | X^{'}z = 0, z \in [-1, 1]^n\}.$$

By setting $\eta = \frac{1}{2}z + \frac{1}{2}e, b = \frac{1}{2}X^{'}e$, it is

$$\min_{\eta}\{y^{'}\eta | X^{'}\eta = b, \eta \in [0, 1]^n\}.$$

For quantile regression $\min_\beta \sum \rho_\tau(y_i - x_i^{'}\beta)$, a similar processing presents the dual formulation:

$$\min_{z}\{y^{'}z | X^{'}z = \tau X^{'}e, z \in [0, 1]^n\}.$$

Since the early 1950's it has been recognized that median regression (L_1 regression) can be formulated as linear programming problems and efficiently solved with some form of the simplex algorithm. For this purpose, the algorithm of Barrodale and Roberts (1974) has proven particularly influential. However, in large statistical applications, the simplex algorithm is regarded as computationally highly demanding. In theory, the worst-case performance of the simplex algorithm shows exponentially increasing number of iterations with sample size.

To solve the L_1 regression for larger data set, several alternate algorithms have been developed. Rather than moving from vertex to vertex around the outer surface of the constraint set as dictated by the simplex, the interior point approach of Karmarkar (1984) solves a sequence of quadratic problems in which the relevant interior of the constraint set is approximated by an ellipsoid. The worst-case performance of the interior point algorithm is demonstrated better than that of the simplex algorithm.

These algorithms developed for the general LP problems may not fully deploy the properties of the original L_1 or quantile regression and have their own shortcomings. For example, the interior point algorithm can only give the approximate solutions of the original problem and rounding has to been done if one requires the same accuracy as that of the simplex algorithm. For some problems, this rounding step requires some significant extra computing time. In this case, some heuristic approaches demonstrate advantages on both speed and accuracy. One of them which will be explored in this paper is the smoothing method.

The smoothing method, also called the continuation method, was used by Clark and Osborne (1986), Madsen and Nielsen (1993) for the L_1 regression. In this paper, we extend this approach to the general quantile regression. The extension is quite natural; however, in quantile regression we consider the solutions for a series of quantiles or the whole regression quantile process. We develop an efficient finite algorithm for computing any finite regression quantiles based on the uniformally finite convergence property of the smoothing algorithm for any finite quantiles.

Numerical comparison shows that the finite smoothing algorithm dominates the simplex algorithm in computing speed. Compared with the interior point algorithm introduced by Portnoy and Koenker (1997), it is competitive overall; however, it is significantly faster than the interior point algorithm when the design matrix in the quantile regression is fat and dense. Section 2 introduces the Huber-type smoothing function and its relation to the objective function of the quantile regression. Section 3 proves the uniformally finite convergence property for the solutions of the smoothed problems. Section 4 describes an efficient algorithm based on the uniformally finite convergence property. Section 5 displays the numerical comparison of this algorithm with the simplex and interior point algorithms.

Bibliography

[1] Barrodale, I. and Roberts, F.D.K. (1973), "An Improved Algorithm for Discrete l_1 Linear Approximation", *SIAM J. Numer. Anal.,* **10**, 839-848.

[2] Clark D. I. and Osborne, M. R. (1986), "Finite Algorithms for Huber's M-estimator", SIAM J. Sci. Statist. Comput., **6**, 72-85.

[3] Chen, C. (2001), "A Finite Smoothing Algorithm for Quantile Regression", Manuscript, SAS Institute Inc.

[4] Huber, P.J. (1981), *Robust Statistics.* John Wiley & Sons.

[5] Karmarkar, N. (1984), "A New Polynomial-time Algorithm for Linear Programming", Combinatorica, **4**, 373-395.

[6] Koenker, R. and Bassett, JR. G. (1978), "Regression Quantiles", Econometrica, Vol. **46**, 33-50.

[7] Madsen, K. and Nielsen, H. B. (1993), "A Finite Smoothing Algorithm for Linear L_1 Estimation", SIAM J. Optimization, Vol **3**, 223-235.

[8] Portnoy, S. and Koenker, R. (1997), "The Gaussian Hare and the Laplacian Tortoise: Computation of Squared-error vs. Absolute-error Estimators", *Statistical Science*, Vol **12**, 279-300.