Simplex Techniques for Quantile Regression Model Selection

Yonggang Yao
SAS Institute Inc.
8/01/2010

Nonparametric Statistics — 2010 JSM, Vancouver, Canada
Outline

- Background
  - Quantile Regression
  - Linear Programming
- Model Selection and Simplex Tableau
  - Greedy Methods
  - Penalty Methods
  - Resampling
- Some Computing Issues
• Three Formulations of Quantile Regression (QR) at level $\tau$

$$
\min \frac{1}{n} \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i \beta)
$$

$$
\min \left( \tau - \frac{1}{2} \right) (\bar{y} - \bar{x} \beta) + \frac{1}{2n} \sum_{i=1}^{n} |y_i - x_i \beta|
$$

$$
\max \sum_{i=1}^{n} y_i a_i \quad \text{s.t.} \quad X'a = (1 - \tau)X'1_n \quad \text{and} \quad a \in [0,1]^n
$$

where $\rho_{\tau}(t) = \tau t^+ + (1 - \tau) t^-$.

• Linear Programming (LP)

$$
\min \ c'z
$$

$$
\text{s.t.} \quad Az = b \quad \text{and} \quad z \geq 0
$$
• Cast QR problem as LP problem

\[
\begin{align*}
    z &= (\beta^+ \quad \beta^- \quad \xi^+ \quad \xi^-)' \\
    c &= (0 \quad 0 \quad \tau/n \quad (1-\tau)/n)' \\
    A &= (X \quad -X \quad I \quad -I) \\
    b &= Y
\end{align*}
\]

LP (Standard Form)
\[
\begin{align*}
    \text{min } c'z \\
    \text{s.t. } Az = b \text{ and } z \geq 0
\end{align*}
\]

where \( c \) and \( z \) are m-vectors with \( m=2p+2n \), \( A \) is a \( n\)-by-

\( m \) matrix, and \( \xi = Y - X\beta \) is the residual vector.
**Simplex Theory**

- Let $B \equiv \{B_1, \ldots, B_n\} \subset \{1,2,\ldots,n\}$ denote an index set.  
  $A_B \equiv [A_{B_1}, \ldots, A_{B_n}]$ denote an invertible sub-matrix of $A$.  
  $z^* \equiv [z_1^*, \ldots, z_m^*]$ is called a basic solution if $z^*$ satisfies:

  $\begin{align*}
  z^*_B &= A_B^{-1} b \\
  z^*_j &= 0 \text{ for } j \in \{1,2,\ldots,m\} \setminus B
  \end{align*}$

- $z^*$ is an optimal solution if

  $\begin{align*}
  z^*_B &= A_B^{-1} b \geq 0 \\
  c - A' A_B^{-1} c_B &\geq 0
  \end{align*}$

- Simplex Tableau

\[
\begin{bmatrix}
-c_B' z_B & c' - c_B' A_B^{-1} A \\
A_B^{-1} b & A_B^{-1} A
\end{bmatrix}
\]

[LP (Standard Form)]

$$\begin{align*}
\text{min} & \quad c'z \\
\text{s.t.} & \quad Az = b \text{ and } z \geq 0
\end{align*}$$
Model Selection and Simplex Tableau

- Cast QR model selection problem as LP problem
  \[
  z = \begin{pmatrix} \beta^+ & \beta^- & \xi^+ & \xi^- & 0 & 0 \end{pmatrix}'
  \]
  \[
  c = \begin{pmatrix} 0 & 0 & \tau/n & (1-\tau)/n & 0 & 0 \end{pmatrix}'
  \]
  \[
  A = \begin{pmatrix} X_1 & -X_1 & I & -I \end{pmatrix} \quad A^* = (X_2 - X_2)
  \]
  \[
  b = Y
  \]
  where $\beta_2$ is forced to be a zero vector.

- Simplex Tableau for Model Selection
  \[
  \begin{bmatrix}
  -c_B'z_B & c' - c_B' A_B^{-1} A & c' - c_B' A_B^{-1} A^* \\
  A_B^{-1} b & A_B^{-1} A & A_B^{-1} A^*
  \end{bmatrix}
  \]
  where $A$ is for $X_1$, and $A^*$ is for $X_2$.

- Use index set $B = \{B_1, \cdots, B_n\}$ to control a model selection process.

LP (Standard Form)
min $c'z$
s.t. $Az = b$ and $z \geq 0$
A model is like a port.
The simplex tableau is like a cargo ship.

Evaluate models with data.
Model Selection and Simplex Tableau

- **Greedy Methods** (Forward, Backward, Stepwise)

- **Fit Criteria**

\[ R^1(\tau) \equiv 1 - \frac{MWAR_F(\tau)}{MWAR_R(\tau)} \quad \text{(vs. } R^2) \]

\[ AIC(\tau) \equiv 2n \log(MWAR(\tau)) + 2p \]

\[ SIC(\tau) \equiv 2n \log(MWAR(\tau)) + p \log n \]

\[ AICC(\tau) \equiv 2n \log(MWAR(\tau)) + 2(p + 1) \frac{n}{n - p - 2} \]

Sawa's \( BIC(\tau) \equiv 2n \log(MWAR(\tau)) + n \log \frac{n + p}{n - p - 2} \)

- Likelihood ratio
- Wald score
- Rank score

\[
\begin{align*}
MWAR_F(\tau) &\equiv \min_{\text{full-model}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho_\tau [y_i - x_i \beta_\tau] \right\}, \\
MWAR_R(\tau) &\equiv \min_{\text{reduced-model}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho_\tau [y_i - x_i \beta_\tau] \right\}.
\end{align*}
\]
Simulation

True Model:  \( p=20 \) and \( n=1000 \)

\[
y = x_1 + 2x_2 + 3x_{10} + x_{12} + x_{15} + x_{18} + e
\]

\[
X = (x_1, \ldots, x_{10}) \sim Unif(0,1) \quad i.i.d.
\]

\[
e \sim N(0,25) \quad i.i.d.
\]
**Simulation**

Forward Selection Summary at quantile level 0.5

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>Objective</th>
<th>p-value (Wald Scores)</th>
<th>QRR</th>
<th>ADJQRR</th>
<th>AIC</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>step entered</td>
<td>function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Intercept</td>
<td>0.875704</td>
<td>0.000000</td>
<td>-0.000500</td>
<td>-131.727</td>
<td>-125.819</td>
</tr>
<tr>
<td>1</td>
<td>x10</td>
<td>0.858893</td>
<td>0.00001</td>
<td>0.019198</td>
<td>0.018707</td>
<td>-151.111</td>
</tr>
<tr>
<td>2</td>
<td>x15</td>
<td>0.850475</td>
<td>0.00001</td>
<td>0.028811</td>
<td>0.027838</td>
<td>-159.960</td>
</tr>
<tr>
<td>3</td>
<td>x12</td>
<td>0.842748</td>
<td>0.00001</td>
<td>0.037634</td>
<td>0.036187</td>
<td>-168.087</td>
</tr>
<tr>
<td>4</td>
<td>x1</td>
<td>0.837204</td>
<td>0.00006</td>
<td>0.043965</td>
<td>0.042047</td>
<td>-173.687</td>
</tr>
<tr>
<td>5</td>
<td>x5</td>
<td>0.832609</td>
<td>0.00058</td>
<td>0.049213</td>
<td>0.046827</td>
<td>-178.192</td>
</tr>
<tr>
<td>6</td>
<td>x18</td>
<td>0.830066</td>
<td>0.00138</td>
<td>0.052117</td>
<td>0.049260</td>
<td>-180.250</td>
</tr>
</tbody>
</table>

* Optimal Value Of Criterion
Selection stopped as the candidate for entry has p-value > 0.1.

Stop Details

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Candidate</th>
<th>Compare For</th>
<th>Effect</th>
<th>Significance</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry</td>
<td>x3</td>
<td>0.13957</td>
<td>&gt;</td>
<td>0.1000</td>
<td>(p-value on Wald Score)</td>
</tr>
</tbody>
</table>
### Simulation

Effects: Intercept x10 x15 x12 x1 x5 x18

Parameter Estimates at quantile level 0.5

| Parameter | DF | Estimate | Error   | Limits      | t Value | Pr > |t|     |
|-----------|----|----------|---------|-------------|---------|------|------|
| Intercept | 1  | -0.0519  | 0.3907  | -0.8186     | -0.13   | 0.8944|
| x10       | 1  | 1.3327   | 0.3115  | 0.7215      | 4.28    | <.0001|
| x15       | 1  | 1.2865   | 0.3086  | 0.6809      | 4.17    | <.0001|
| x12       | 1  | 1.1219   | 0.3066  | 0.5203      | 3.66    | 0.0003|
| x1        | 1  | 0.8864   | 0.3010  | 0.2957      | 2.94    | 0.0033|
| x5        | 1  | 0.7483   | 0.3030  | 0.1536      | 2.47    | 0.0137|
| x18       | 1  | 0.6918   | 0.3021  | 0.0990      | 2.29    | 0.0222|
Model Selection and Simplex Tableau

• Penalty Methods
  LASSO penalty, OSCAR penalty, Grouped LASSO penalty

• Manipulating Simplex Algorithm for Penalty Methods
  For example, LASSO penalty can be measured by using $a$ vector as follow:

$$\begin{bmatrix}
-c_B'z_B & c' B^{-1}_B A \\
-a_B' z_B & a' B^{-1}_B A \\
B^{-1}_B b & B^{-1}_B A \\
\end{bmatrix}$$

LP (Parametric - Cost Form)
$$\min \ c'z + a'z$$
$$\text{s.t. } Az = b \text{ and } z \geq 0$$

where $a = (1, 1, 0, 0)$ according to $z = (\beta^+, \beta^-, \xi^+, \xi^-)'$. 
Simulation

True Model: \( p = 11 \) and \( n = 1000 \).

\[
y = \sum_{g=1}^{3} \sum_{i=1}^{p_g} x_{gi} \beta_{gi} + e
\]

True \( \beta = ((2,3,2),(0,0,0),(0,0,0),(-3,2,-2)) \)

\( X = (x_1, \ldots, x_{10}) \sim N(0,1) \) i.i.d.

\( e \sim N(0,50) \) i.i.d.
Solution Path for LASSO QR

Penalty: \( s = \sum_{i=1}^{p} |\beta_i| \).

True \( \beta = ((2,3,2),(0,0,0,0),(-3,2,-2)) \)
True $\beta = ((2,3,2),(0,0,0,0),(-3,2,-2))$
Solution Path for Grouped-LASSO QR

Penalty: \( s = \sum_{g=1}^{G} \max \left( |\beta_{g1}|, \ldots, |\beta_{gg_p}| \right). \)

True \( \beta = ((2,3,2),(0,0,0,0),(-3,2,-2)) \)
Applications of Simplex Techniques

• Resampling
  Cross-validation, Bootstrap

• Manipulating Simplex Algorithm for Resampling
  1. Check whether an observation is active for a fitted model.
  2. Drive-out some observations by changing the objective function.
Simplex Tableau can be used to:

• update an optimal partial model to another optimal partial model or full model.

• add extra constraints on a model.

• update an optimal model on a subset of a dataset to the optimal model on another subset of the dataset.
Computational Goals

- High-performance computing
- Massive data processing
- Re-usable programs
• Parallel Computation can expedite Tableau Simplex algorithm on

1. Building initial tableau
2. Sorting/Ordering positive tableau rows
3. Changing the signs of tableau rows
4. Pivot Updating


