MM algorithm for Quantile regression for censored data with missing observations

Samidha Shetty

Pennsylvania State University

4^{th} December 2018

Quantile regression

A linear quantile regression model is given as below :

$$Y_i = Z_i^T \theta_q + r_i, \quad i = 1, \dots, n$$

such that $P(r_i \leq 0 | Z_i) = q$ or $E\{q - I(r_i \leq 0) | Z_i)\} = 0$
Here Y_i : response, Z_i : vector of covariates, θ_q : unknown coefficient vector (which depends on q), r_i : error term

What are the advantages of quantile regression over mean regression ?

Koenker and Bassett(1978) defined $\hat{\theta}$ as the minimizer of

$$L(\theta) = \sum_{i=1}^{n} \rho_q[y_i - z_i^T \theta] = \sum_{i=1}^{n} \rho_q(r_i(\theta))$$

where $\rho_q(r) = |r|[q - I(r \leq 0)]$.

But this is not easily optimized as $\rho_q(r)$ is non-differentiable at $r = 0$.

MM algorithm

Hunter and Lange (2000) introduced an MM algorithm to optimize $L(\theta)$.

First, $L(\theta)$ is approximated by $L_{\varepsilon}(\theta) = \sum_{i=1}^{n} \rho_q^{\varepsilon}(r_i)$, where,

$$\rho_q^{\varepsilon}(r) = \rho_q(r) - \frac{\varepsilon}{2} ln(\varepsilon + |r|)$$

Second, the approximated function is minimized using an MM algorithm.
At k^{th} iteration $\rho_q^{\varepsilon}(r)$ is majorized by

$$\zeta_q^{\varepsilon}(r|r^k) = \frac{1}{4}\left[\frac{(r)^2}{\varepsilon + |r^k|} + (4q-2)r + c\right]$$

where is c is such that $\zeta_q^{\varepsilon}(r^k|r^k) = \rho_q^{\varepsilon}(r^k)$.

MM algorithm

Thus the majorizer for $L_\varepsilon(\theta)$ is given as

$$Q_\varepsilon(\theta|\theta^k) = \sum_{i=1}^{n} \zeta_q^\varepsilon(r_i|r_i^k)$$

In the linear case, one can solve explicitly for θ^{k+1}, but otherwise, just reducing the value of $Q_\varepsilon(\theta|\theta^k)$ at each iteration suffices.

MM algorithm for Quantile regression

1. Initialize θ^0 and small constant ε such that $\varepsilon n|\ln\varepsilon| = \tau$. Set $k = 0$.
2. At every k^{th} iteration $\theta^{k+1} = \theta^k + \alpha^k\phi_\varepsilon^k$ where α^k is step size and ϕ_ε^k is step direction.
3. Replace $k = k + 1$. Until $\frac{Q_\varepsilon(\theta^{k+1}|\theta^k) - Q_\varepsilon(\theta^k|\theta^k)}{Q_\varepsilon(\theta^k|\theta^k)} < \tau$.

Censored data

Censoring, roughly speaking, is when the value of a observation is only partially known.
Right censoring : We don't have the actual value of the observation, but instead know that it is above a certain value i.e. we observe $Y_i = min(T_i, c_i)$ and $\Delta_i : I(T_i \leq c_i)$ is an indicator of censoring.
Xie et al. (2015) used an inverse probability weighted estimating function of the form

$$\sum_{i=1}^{n} \frac{\Delta_i}{G(y_i|Z_i)} \rho_q[y_i - z_i^T \theta]$$

where $G(Y_i|Z_i)$ is the survival function which is estimated using the Kaplan-Meier estimator. When c_i is independent of covariates,

$$\widehat{G}(t|Z_i) = \widehat{G}(t) = \prod_{s \leq t} \left\{ 1 - \frac{\#\text{of deaths before time s}}{\#\text{of surviving people at time s}} \right\}$$

Missing values

To deal with missing values in quantile estimation (Chen et al. (2014) devised an inverse probability weighting method that estimates the probability weights non-parametrically.
The objective function is modifed as follows :

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{p}(X_i)} \rho_q[y_i - z_i^T \theta]$$

where X_i is the matrix of response and subset of covariates which have complete data, δ_i is indicator of completeness of data for i^{th} observation.
$\widehat{p}(X_i) = \frac{\sum_{j=1}^{n} K_h(X_i - X_j)\delta_j}{\sum_{j=1}^{n} K_h(X_i - X_j)}$ where $K_h(u) = K(u/h)/h^d$. Here d is the dimension of X_i and $K(\cdot)$ is a d-variate probability density function.
I have combined these two methods, and get the following objective function :

$$\sum_{i=1}^{n} \frac{\Delta_i}{\widehat{G}(y_i|Z_i)} \frac{\delta_i}{\widehat{p}(X_i)} \rho_q[y_i - z_i^T \theta]$$

Simulation Study

Model : $Y_i = 4.5Z_{1i} - 2Z_{2i} + r_i$
where $Z_{1i} \sim Normal(0, 1)$, $Z_{2i} \sim Uniform(-3, 3)$ and $r_i \sim Normal(0, 1)$. I have used quantiles of Y to censor the data to attain the particular amount of censoring. Only covariate Z_2 has missing values.

For θ_1 :

C %	M %	$q = 0.25$		$q = 0.5$		$q = 0.75$	
		Bias	MSE	Bias	MSE	Bias	MSE
25%	30%	-0.2486	0.0719	-0.2474	0.0693	-1.0631	1.1749
50%	30%	-0.2172	0.0586	-2.2552	5.1272	-2.2495	5.1068
25%	50%	-0.2387	0.0703	-0.6786	0.5033	-1.0252	1.1038

For θ_2 :

C %	M %	$q = 0.25$		$q = 0.5$		$q = 0.75$	
		Bias	MSE	Bias	MSE	Bias	MSE
25%	30%	0.1140	0.0155	0.1136	0.0153	0.4406	0.2009
50%	30%	0.1167	0.0163	0.9935	0.9949	0.9829	0.9738
25%	50%	0.1058	0.0145	0.2982	0.0973	0.4386	0.2009

Boxplot for θ_1 estimates

Situation 1 : 25% right censoring and 30% missing observations
Situation 2 : 50% right censoring and 30% missing observations
Situation 3 : 25% right censoring and 50% missing observations

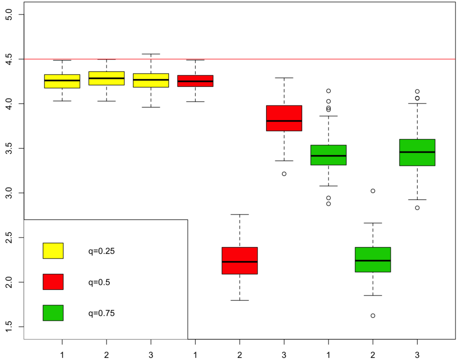

FIGURE – Box plot for θ_1 estimates

Observations and future work

Observations and challenges

- The method estimates better for smaller quantiles.
- Increase in censoring affects the estimates drastically.
- The estimates for θ_1 are slightly worse than those for θ_2.
- There seems to be a bias present in the estimation.
- Computation time : For n=500 it took 23.6 sec, n=1000 it took 130 sec and n=2000 it took 835 sec.

Future work

- Derive theoretical results for the combination of the two methods.
- Considering case where censoring is covariate dependent.
- Extending the methods to partially linear quantile regression.