Bayesian Quantile Regression using a Mixture of Polya Trees

Minzhao Liu and Mike Daniels

Department of Statistics
University of Florida

2012.4.4

Outline

Quantile Regression

Polya Tree

QR with Polya Tree

What is Quantile Regression

Suppose random variable Y has cdf F, then the τ-th quantile of Y is

$$Q_Y(\tau) = \inf \{y : F(y) \geq \tau\},$$

Furthermore, if we have covariates $\mathbf{X}$, then the quantile regression parameter $\boldsymbol{\beta}$ satisfies this condition

$$Q_Y(\tau|x) = x'\boldsymbol{\beta}(\tau),$$

if F is continues , then

$$\Pr(y \leq x'\boldsymbol{\beta}(\tau)) = \tau.$$

Why We Need Quantile Regression

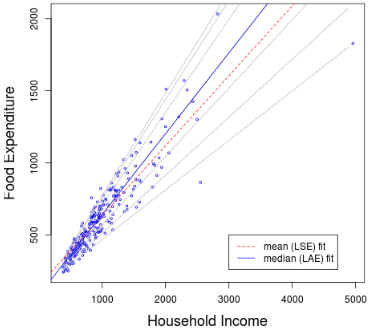


Figure: Engel Curves for Food: This figure plots data taken from Engel's (1857) study of the dependence of households' food expenditure on household income.

Why (Cont'd)

Advantage over regular mean regression:

- more information provided,
- slope may vary with different τ.

How to Get Estimates (Koenker)

Recall least squared estimator for mean regression model is :

$$\boldsymbol{\beta} = \underset{b \in \mathrm{R}^p}{\arg\min} \sum_{i=1}^{n} (y_i - x_i' b)^2$$

In quantile regression , we have : The true τ-th quantile regression parameter $\boldsymbol{\beta}(\tau)$ minimizes the expectation of the check function:

$$\boldsymbol{\beta}(\tau) = \underset{b \in \mathrm{R}^p}{\arg\min}\, \mathrm{E}\left(\rho_\tau(y - x_i' b)\right),$$
$$\rho_\tau(z) = z\left(\tau - I(z < 0)\right).$$

Empirically, we take

$$\boldsymbol{\beta}(\tau) = \underset{b \in \mathrm{R}^p}{\arg\min} \sum_{i=1}^{n} \left(\rho_\tau(y - x_i' b)\right),$$

Equivalence of ASL and Check Function (Yu-2001)

Asymmetric Laplace Distribution pdf:

$$f_\tau(z;\mu,\sigma) = \frac{\tau(1-\tau)}{\sigma}\exp\left\{-\rho_\tau\left(\frac{z-\mu}{\sigma}\right)\right\}$$

Property:

$$F_z(\mu) = \tau$$

If we assume

$$\begin{aligned} y &= x\boldsymbol{\beta} + \epsilon \\ \epsilon &\sim \mathrm{ASL}_\tau(0,\sigma) \end{aligned}$$

then,

$$L(\boldsymbol{\beta},\sigma|\mathbf{Y}) \propto \sigma^{-n}\exp\left\{-\sum_{i=1}^{n}\rho_\tau\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\sigma}\right)\right\}$$

$$\Leftrightarrow \quad \boldsymbol{\beta}(\tau) = \arg\min_{b\in\mathrm{R}^p}\sum_{i=1}^{n}\left(\rho_\tau(y - x_i'b)\right),$$

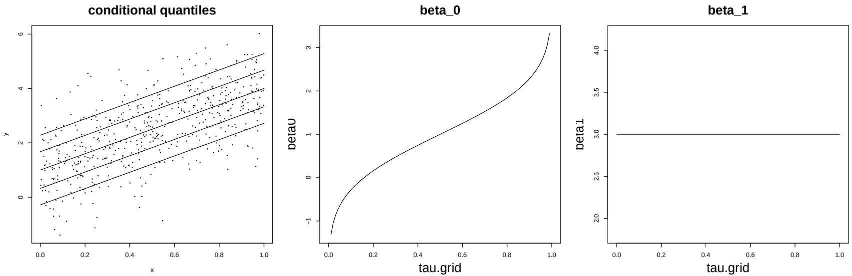
Remarks

- no distribution assigned in check function
- one ALD distribution for all models
- for different τ, separate model needed

Our Proposal

Consider location-shift model

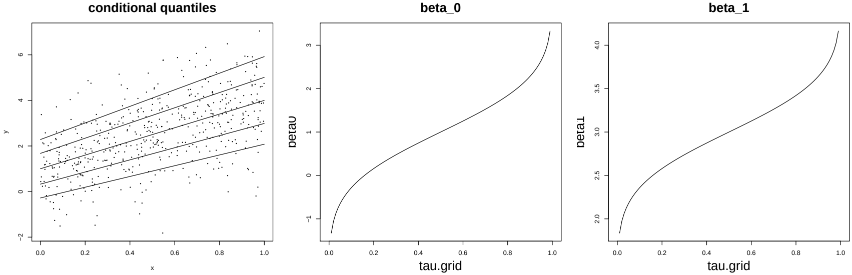
$$Y = X\boldsymbol{\beta} + \epsilon, \quad \epsilon \sim F_\epsilon$$
$$\Rightarrow \boldsymbol{\beta}(\tau) = \boldsymbol{\beta} + F_\epsilon^{-1}(\tau)\mathbf{e}_1$$



$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i.$$

Heterogeneity

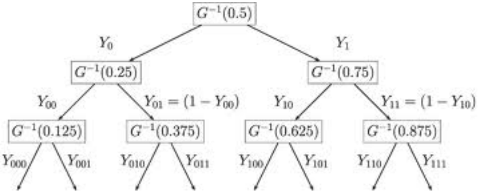
$$Y = x\boldsymbol{\beta} + (x\boldsymbol{\gamma})\epsilon, \quad \epsilon \sim F_\epsilon$$
$$\Rightarrow \boldsymbol{\beta}(\tau) = \boldsymbol{\beta} + \boldsymbol{\gamma} F_\epsilon^{-1}(\tau)$$



$$y_i = \beta_0 + \beta_1 x_{i1} + (1 + x_{i1})\epsilon_i$$

Question: How to get $F_\epsilon^{-1}(\tau)$,

A Flexible Nonparametric Prior: Polya Tree



$$G|\Pi, \mathcal{A} \sim PT(\Pi, \mathcal{A})$$

PT parameters:

1. Partitions: $\Sigma = \{B_0, B_1, B_{00}, \ldots\}$
2. Weights $\mathcal{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_\epsilon, \ldots\}$
3. $Y_{\epsilon 0} \sim Beta(\alpha_{\epsilon 0}, \alpha_{\epsilon 1})$
4. $\alpha_{\epsilon 0} = \alpha_{\epsilon 1} = cj^2$, absolutely continuous
5. Partial / finite PT , M level

Why and advantages

1. flexible, nonparametric
2. posterior tractable

$$
\begin{aligned}
x|G &\sim G \\
G &\sim PT(B, \mathcal{A}) \\
\Rightarrow G|x &\sim PT(B, \mathcal{A}^*) \quad \text{with} \quad \alpha^*_\epsilon = \begin{cases} \alpha_\epsilon + 1 & \text{if } x \in B_\epsilon \\ \alpha_\epsilon & \text{otherwise} \end{cases}
\end{aligned}
$$

3. easy to get posterior quantile
4. easy to fix median or other quantiles
5. can be absolute ly continuous , while DP can not

Mixture of Polya Trees

Since the behavior of polya tree prior highly depends on the partition parameter Π, thus a mixture of Polya tree prior was defined this way:

$$
\begin{aligned}
x|G &\sim G \\
G &\sim PT(B, \mathcal{A}) \\
\Rightarrow G|\Pi^{\theta}, \mathcal{A}^{\theta} &\sim PT(\Sigma^{\mu,\sigma}, \mathcal{A}^{c})
\end{aligned}
$$

where Σ partition was constructed based on baseline measure $\mathrm{N}(\mu, \sigma^2)$. and then assign $\theta = (\mu, \sigma^2, c)$ a prior.

$$\theta \sim \pi(\theta)$$

Predictive Density, Cumulative, and Quantiles

$$X_1, \ldots, X_n | G \sim G$$
$$G | \Pi^{\sigma^2}, \mathcal{A}^c \sim PT(\Pi^{\sigma^2}, \mathcal{A}^c)$$

where $G_0 = \mathrm{N}(0, \sigma^2)$ is the baseline measure, and $g_0(x)$ is its density function. Then, from the partial PT, the predictive density function is

$$f(x | X_1, \ldots, X_n) = \left[\prod_{j=2}^{M} \frac{cj^2 + n(\epsilon(j, x) | \mathbf{X})}{2cj^2 + n(\epsilon(j-1, x) | \mathbf{X})} \right] 2^{M-1} g_0(x)$$

Predictive Cumulative, and Quantiles (Cont'd)

Integrating the predictive density function,

$$F(x|X_1,\ldots,X_n) = \sum_{i=1}^{N-1} P_i + P_N\left(G_0(x)2^M - (N-1)\right)$$

$$P_i = \frac{1}{2}\prod_{j=2}^{M}\frac{cj^2 + n(\epsilon(j,x)|\mathbf{X})}{2cj^2 + n(\epsilon(j-1,x)|\mathbf{X})}$$

$$N = \left\lceil 2^M G_0(x) + 1\right\rceil$$

By inverting the predictive cumulative density function,

$$F^{-1}_{X|\mathbf{X}}(\tau) = G_0^{-1}\left[\frac{\tau - \sum_{i=1}^{N} P_i + N\cdot P_N}{2^M P_N}\right]$$

$$N \quad \text{satisfy} \quad \sum_{i=1}^{N-1} P_i < \tau \leq \sum_{i=1}^{N} P_i$$

Quantile Regression with PT

Suppose we have a location-shift model, with heterogeneity form:

$$
\begin{aligned}
& y_i = x_i\boldsymbol{\beta} + (x_i\boldsymbol{\gamma})\epsilon_i \\
& \epsilon_i | G \sim G \\
& G | \Pi, \mu, \sigma, \mathcal{A} \sim PT(\Pi^{\mu,\sigma}, \mathcal{A})
\end{aligned}
$$

in order not to confound with location parameter $\boldsymbol{\beta}$, we fix ϵ's median as 0 ($\mu = 0$). Also, to deal with the identifiability problem of $\boldsymbol{\gamma}$, the first element of $\boldsymbol{\gamma}$ is set to be 1.
Estimation:

$$
\begin{aligned}
\boldsymbol{\beta}(\tau) &= \boldsymbol{\beta} + \boldsymbol{\gamma} F_{\epsilon}^{-1}(\tau) \\
p(\boldsymbol{\beta}(\tau)|\mathbf{Y}) &= p(h(\boldsymbol{\beta}, \boldsymbol{\gamma}, F_{\epsilon}^{-1})|\mathbf{Y})
\end{aligned}
$$

Simulation and Comparison

Compare our quantile regression model with Polya trees priors (HeterPTlm) with Koenker's 'rq' function in R package 'quantreg'.

- $Y_i = 1 + x_{i1} + x_{i2} + \mathrm{N}(0,1)$
- $Y_i = 1 + x_{i1} + x_{i2} + \mathrm{Gamma}(3,1)$
- $Y_i = 1 + x_{i1} + x_{i2} + \mathrm{MixtureNormal}$
- $Y_i = 1 + x_{i1} + x_{i2} + (1 - 0.5x_{i1} + 0.5x_{i2})\mathrm{MixtureNormal}$
- $Y_i = 1 + x_{i1} + x_{i2} + (1 - 0.5x_{i1} + 0.5x_{i2})\mathrm{Gamma}(3,1)$

where Mixture Normal $\sim 0.5\mathrm{N}(-2,1) + 0.5\mathrm{N}(2,1)$.
100 datasets and each dataset contains 100 observations.
Evaluate $\tau = 0.5$ and $\tau = 0.9$.

Results

Model	MSE rq	HeterPTlm	Coverage rq	HeterPTlm
M1.5	1.39(0.13)	1.06(0.12)	0.82	0.93
M2.5	3.31(0.36)	3.18(0.35)	0.89	0.95
M3.5	17.2(1.48)	2.21(0.30)	0.85	0.95
M4.5	95.4(6.69)	9.87(1.09)	0.86	0.95
M5.5	14.3(1.36)	10.0(1.04)	0.89	0.9
M1.9	2.35(0.26)	2.07(0.23)	0.95	0.96
M2.9	15.2(1.51)	13.5(1.61)	0.9	0.89
M3.9	3.68(0.46)	3.94(0.56)	0.93	0.97
M4.9	25.1(2.66)	14.8(1.41)	0.88	0.94
M5.9	73.7(8.48)	60.2(8.14)	0.93	0.87

length of credible interval from HeterPTlm is also shorter than that from rq function.