High dimensional Bayesian quantile regression

Prithwish Bhaumik
The University of Texas at Austin
and
Lizhen Lin
The University of Texas at Austin

Basic set up

- Random sample $Y_1, Y_2, \ldots, Y_n$.

Basic set up

- Random sample $Y_1, Y_2, \ldots, Y_n$.
- Regression model: $Y_i = \mathbf{X}_i^T \beta + \varepsilon_i$, $i = 1, \ldots, n$.

Basic set up

- Random sample $Y_1, Y_2, \ldots, Y_n$.
- Regression model: $Y_i = \mathbf{X}_i^T \beta + \varepsilon_i$, $i = 1, \ldots, n$.
- $q_{Y_i}(\tau|\mathbf{X}_i) = \mathbf{X}_i^T \beta$, $0 < \tau < 1$.

Basic set up

- Random sample $Y_1, Y_2, \ldots, Y_n$.
- Regression model: $Y_i = \mathbf{X}_i^T \beta + \varepsilon_i$, $i = 1, \ldots, n$.
- $q_{Y_i}(\tau|\mathbf{X}_i) = \mathbf{X}_i^T \beta$, $0 < \tau < 1$.
- $q_{\varepsilon_i}(\tau|\mathbf{X}_i) = 0$.

Basic set up

- Random sample $Y_1, Y_2, \ldots, Y_n$.
- Regression model: $Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \ldots, n$.
- $q_{Y_i}(\tau|\mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\beta}$, $0 < \tau < 1$.
- $q_{\varepsilon_i}(\tau|\mathbf{X}_i) = 0$.
- $\mathbf{X}_i \in \mathbb{R}^q$, $q \geq 1$ for $i = 1, \ldots, n$.

Basic set up

- Random sample $Y_1, Y_2, \ldots, Y_n$.
- Regression model: $Y_i = \mathbf{X}_i^T \beta + \varepsilon_i$, $i = 1, \ldots, n$.
- $q_{Y_i}(\tau | \mathbf{X}_i) = \mathbf{X}_i^T \beta$, $0 < \tau < 1$.
- $q_{\varepsilon_i}(\tau | \mathbf{X}_i) = 0$.
- $\mathbf{X}_i \in \mathbb{R}^q$, $q \geq 1$ for $i = 1, \ldots, n$.
- Objective: infer on β.

Difference with usual regression

Difference with usual regression

- Usual regression models the mean of the random variable.

Difference with usual regression

- Usual regression models the mean of the random variable.
- Only concerned with the measure of central tendency

Difference with usual regression

- Usual regression models the mean of the random variable.
- Only concerned with the measure of central tendency
- quantile regression considers regressing any arbitrary quantile of Y on $\mathbf{X}$.

Difference with usual regression

- Usual regression models the mean of the random variable.
- Only concerned with the measure of central tendency
- quantile regression considers regressing any arbitrary quantile of Y on $\mathbf{X}$.
- more informative about the distribution of the random variable.

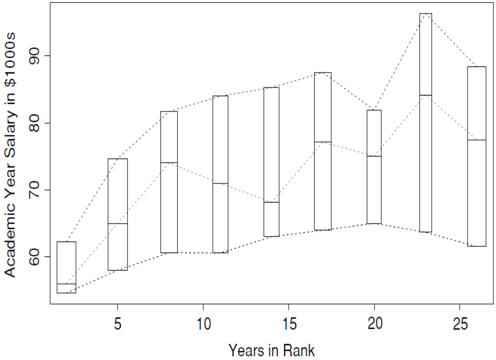

Figure: *1995 ASA academic salary survey for full professors of Statistics in U.S. colleges and universities*

Two frameworks:

Two frameworks:

- q is fixed

Two frameworks:

- q is fixed
- q increases with n: high dimensional set up.

Posterior contraction

Posterior contraction rate r_n^{-1}, $r_n \to \infty$:
There exists $M > 0$ such that

$$\Pi\left(r_n \|\boldsymbol{\theta} - \theta_0\| > M | \mathbf{Y}\right) \xrightarrow{P} 0$$

Bernstein-von Mises Theorem

Bernstein-von Mises Theorem

- In a regular parametric model, Bayesian and frequentist distributions of $\sqrt{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ are nearly equal for large sample sizes and the common distribution is a Gausian distribution with mean zero. Here $\hat{\boldsymbol{\theta}}$ is the corresponding Bayes estimator or the MLE or some other efficient estimator (in most cases).

Bernstein-von Mises Theorem

- In a regular parametric model, Bayesian and frequentist distributions of $\sqrt{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ are nearly equal for large sample sizes and the common distribution is a Gausian distribution with mean zero. Here $\hat{\boldsymbol{\theta}}$ is the corresponding Bayes estimator or the MLE or some other efficient estimator (in most cases).

Bernstein-von Mises Theorem

- In a regular parametric model, Bayesian and frequentist distributions of $\sqrt{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ are nearly equal for large sample sizes and the common distribution is a Gausian distribution with mean zero. Here $\hat{\boldsymbol{\theta}}$ is the corresponding Bayes estimator or the MLE or some other efficient estimator (in most cases).
- This is a great reconciliation of two very different ways of quantifying uncertainties- frequentist and Bayes.

Contribution

Contribution

- Consider a Bayesian quantile regression approach by putting a prior on the coefficients of the regression function.

Contribution

- Consider a Bayesian quantile regression approach by putting a prior on the coefficients of the regression function.
- Establish a Bernstein-von Mises theorem for the posterior distribution of β.

Contribution

- Consider a Bayesian quantile regression approach by putting a prior on the coefficients of the regression function.
- Establish a Bernstein-von Mises theorem for the posterior distribution of $\boldsymbol{\beta}$.
- Posterior contraction rate is $\frac{q(\log q)^{1/2}}{\sqrt{n}}$.

- Proposed model: $Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \ldots, n$.

- Proposed model: $Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \ldots, n$.
- $\mathbf{X}_i \overset{iid}{\sim} G$ with density g.

- Proposed model: $Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \ldots, n$.
- $\mathbf{X}_i \overset{iid}{\sim} G$ with density g.
- $\mathrm{E}(\mathbf{X}_i) = \mathbf{0}$, $\mathrm{E}(\mathbf{X}_i \mathbf{X}_i^T) = \mathbf{I}_q$.

- Proposed model: $Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \ldots, n$.
- $\mathbf{X}_i \overset{iid}{\sim} G$ with density g.
- $\mathrm{E}(\mathbf{X}_i) = \mathbf{0}$, $\mathrm{E}(\mathbf{X}_i \mathbf{X}_i^T) = \mathbf{I}_q$.
- True model: $Y_i = \mathbf{X}_i^T \boldsymbol{\beta}_0 + \varepsilon_i$, $i = 1, \ldots, n$.

- Proposed model: $Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \ldots, n$.
- $\mathbf{X}_i \overset{iid}{\sim} G$ with density g.
- $\mathrm{E}(\mathbf{X}_i) = \mathbf{0}$, $\mathrm{E}(\mathbf{X}_i \mathbf{X}_i^T) = \mathbf{I}_q$.
- True model: $Y_i = \mathbf{X}_i^T \boldsymbol{\beta}_0 + \varepsilon_i$, $i = 1, \ldots, n$.
- Working distribution:

$$f(y, \mathbf{x}|\boldsymbol{\beta}) = \tau(1-\tau)\exp\{-(y - \mathbf{x}^T\boldsymbol{\beta})(\tau - I(y \leq \mathbf{x}^T\boldsymbol{\beta})\}g(\mathbf{x})$$

.

Prior

Prior

- Prior:

$$\pi(\boldsymbol{\beta}|\gamma, \lambda_0, \lambda_1) = \prod_{j=1}^{q} \left[(1-\gamma)\psi(\beta_j|\lambda_0) + \gamma\psi(\beta_j|\lambda_1)\right],$$

where $\psi(\beta_j|\lambda) = \frac{\lambda}{2}\exp\{-\lambda|\beta_j|\}$, $\lambda_0 \gg \lambda_1$.

Let us denote

$$\pi_n^*(\mathbf{u}) = \frac{\pi\left(\beta_0 + \mathbf{u}/\sqrt{n}\right) Z_n(\mathbf{u})}{\int \pi\left(\beta_0 + \mathbf{w}/\sqrt{n}\right) Z_n(\mathbf{w}) d\mathbf{w}}$$

and $\boldsymbol{\Delta}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\beta_0}(Y_i, \mathbf{X}_i)$, where

$$\mathbf{u} = \sqrt{n}(\beta - \beta_0),$$

$$Z_n(\mathbf{u}) = \prod_{i=1}^n \frac{f\left(Y_i, \mathbf{X}_i | \beta_0 + \mathbf{u}/\sqrt{n}\right)}{f\left(Y_i, \mathbf{X}_i | \beta_0\right)}$$

and

$$\dot{\ell}_{\beta_0}(Y, \mathbf{X}) = \mathbf{X}(\tau - I(Y \leq \mathbf{X}^T \beta_0)).$$

Theorem

$$\int \left|\pi_n^*(\mathbf{u}) - \phi_q(\mathbf{u}; \boldsymbol{\Delta}_n, \mathbf{I}_q)\right| d\mathbf{u} \xrightarrow{P} 0,$$

where $\phi_q(; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the pdf of a q-component Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Outline of the proof

The model is differentiable in quadratic mean, that is,

$$\int \left(f^{1/2}(y, \mathbf{x}|\beta_0 + \mathbf{u}/\sqrt{n}) - f^{1/2}(y, \mathbf{x}|\beta_0) - \frac{1}{2\sqrt{n}} \mathbf{u}^T \dot{\ell}_{\beta_0} f^{1/2}(y, \mathbf{x}|\beta_0) \right)^2 dyd\mathbf{x} = O\left(\frac{\|\mathbf{u}\|^3}{n^{3/2}} \right).$$

Let us denote

$$\begin{aligned} p_n &= f(y, \mathbf{x}|\boldsymbol{\beta}_0 + \mathbf{u}/\sqrt{n}) \\ p &= f(y, \mathbf{x}|\boldsymbol{\beta}_0) \end{aligned} \tag{1}$$

Now we have the following lemma.

Lemma (Local asymptotic normality)

(i) For $\|\mathbf{u}\| \lesssim q(\log q)^{1/2}$,

$$\log \prod_{i=1}^{n} \frac{p_n}{p}(Y_i, \mathbf{X}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{u}^T \dot{\ell}_{\boldsymbol{\beta}_0}(Y_i, \mathbf{X}_i) - \frac{1}{2}\|\mathbf{u}\|^2 + O_P\left(\frac{\|\mathbf{u}\|^{2+\delta_1}}{n^{\delta_2}}\right),$$

where $q^{\delta_1}(\log q)^{\delta_1/2}/n^{\delta_2} \to 0$.
(ii) For $\|\mathbf{u}\| \lesssim (q \log q)^{1/2}$,

$$\log \prod_{i=1}^{n} \frac{p_n}{p}(Y_i, \mathbf{X}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{u}^T \dot{\ell}_{\boldsymbol{\beta}_0}(Y_i, \mathbf{X}_i) - \frac{1}{2}\|\mathbf{u}\|^2 + O_P\left(\frac{\|\mathbf{u}\|^{2+\delta_1^*}}{n^{\delta_2^*}}\right),$$

where $q^{\delta_1^*}(\log q)^{\delta_1^*/2}/n^{\delta_2^*} \to 0$.

Let us denote $\tilde{Z}_n(\mathbf{u}) = \exp[\mathbf{u}^T \boldsymbol{\Delta}_n - \|\mathbf{u}\|^2/2]$ and $\lambda_n^* = (q \log q)^{\delta_1^*/2}/n^{\delta_2^*}$.

Lemma

For any $C > 0$, there exists $B' > 0$ such that for all sufficiently large n, with any preassigned large probability

$$\left(\int \tilde{Z}_n(\mathbf{u}) d\mathbf{u} \right)^{-1} \int_{\|\mathbf{u}\| \le C(q \log q)^{1/2}} \left| Z_n(\mathbf{u}) - \tilde{Z}_n(\mathbf{u}) \right| d\mathbf{u} \le B' q \lambda_n^*.$$

Lemma

There exists $B_0, \varepsilon_1 > 0$ *such that*

$$\mathrm{E}\left|Z^{1/2}(\mathbf{u}_1) - Z^{1/2}(\mathbf{u}_2)\right|^2 \leq B_0\|\mathbf{u}_1 - \mathbf{u}_2\|^2, \;\; \mathbf{u}_1,\, u_2 \in n^{1/2}(\beta - \beta_0)$$

and

$$\mathrm{E}Z_n^{1/2}(\mathbf{u}) \leq \exp\left(-\varepsilon_1\|\mathbf{u}\|^2\right).$$

Lemma

For any $0 < \delta < 1$,

$$P\left\{\int Z_n(\mathbf{u})\pi\left(\boldsymbol{\beta}_0 + \frac{\mathbf{u}}{\sqrt{n}}\right) d\mathbf{u} < \pi(\boldsymbol{\beta}_0)\frac{\delta^q}{4}\right\} \leq 4B_0^{1/2}\delta.$$

Lemma (Posterior consistency)

There exists $C > 0$ *such that*

$$\mathrm{E}\left(\int_{\|\mathbf{u}\|>Cq(\log q)^{1/2}} \pi_n^*(\mathbf{u})d\mathbf{u}\right) \to 0 \; as \; n \to \infty$$

.

Lemma

For any $C_2, c > 0$, we can find $B_2, C_1 > 0$ such that with probability approaching 1,

$$\int_{C_1(q\log q)^{1/2}\le\|\mathbf{u}\|\le C_2 q(\log q)^{1/2}} Z_n(\mathbf{u})d\mathbf{u} \le B_2\exp[-cq\log q].$$

Lemma (Estimate of tail probability)

For any $c > 0$, there exists $C > 0$ such that with any preassigned probability,

$$\int_{\|\mathbf{u}\| > Cq^{1/2}} \phi_q(\mathbf{u}; \boldsymbol{\Delta}_n, \mathbf{I}_q) \leq \exp[-cq].$$