Distributed Inference for Quantile Regression Processes

Shih-Kang Chao
Department of Statistics
University of Missouri

June, 2019
ICSA Applied Statistics Symposium

Modern applications lead to data sets so large that cannot be stored in a single machine

- Social media (views, likes, comments, images...)
- Meteorological and environmental surveillance
- Transactions in e-commerce
- Others...

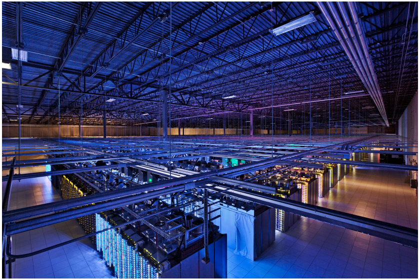


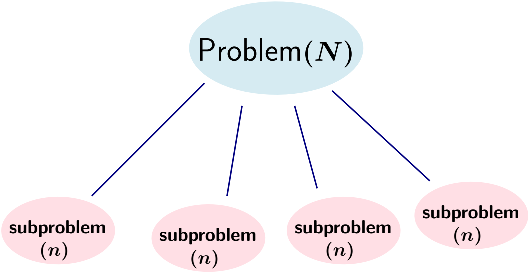
Figure: A Google server room in Council Bluffs, Iowa.

Computational bottlenecks

- Big data cannot fit into the memory of typical computers
  - Many classical statistical methods cannot be performed, e.g. maximum likelihood, Bayesian analysis...
- Buying a computer with huge memory is expensive
- Common solution: buying many usual computers

Divide and Conquer (D&C) framework

- Divide N data into m subsamples. $n = N/m$: subsample size
- Each local machine processes one subsample
- Central computer aggregates outcomes from local machines (costs computational overhead)
- Applied with communication-efficient algorithm: minimize the number of times the central computer calls local machines



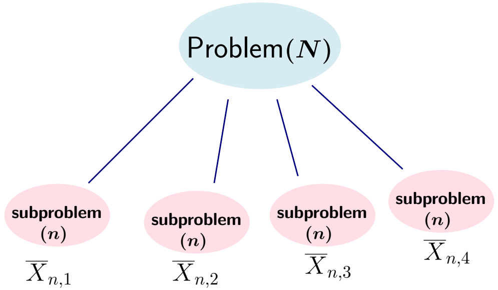
D&C looks nice...but is it always accurate?

I will give two examples

1st example: sample mean ✓

2nd example: sample quantile ?

Example 1: sample mean

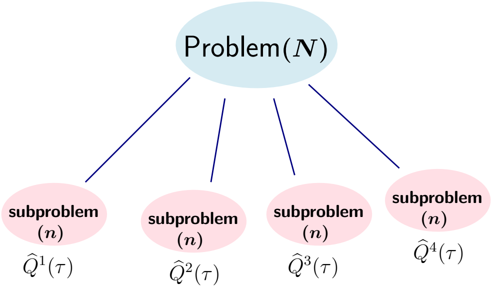


$$\underbrace{\frac{1}{4}\sum_{s=1}^{4}\overline{X}_s}_{\text{Avg. local means}} = \frac{1}{4n}\sum_{s=1}^{4}\sum_{i=1}^{n} X_{is} = \frac{1}{N}\sum_{i=1}^{N} X_i = \underbrace{\overline{X}_N}_{\text{global mean}}.$$

It fits!

Example 2: sample quantile at $\tau \in (0,1)$

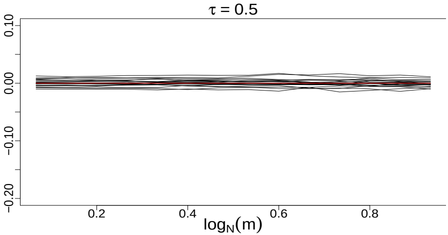
Quantile $\widehat{Q}(\tau) = \lfloor N\tau \rfloor$ order statistics

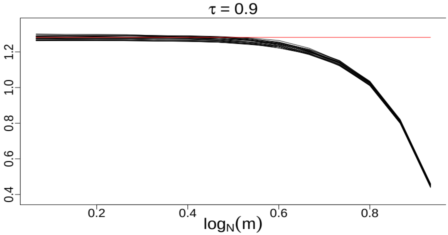


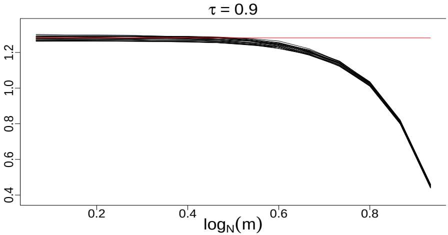
$$\underbrace{\frac{1}{4}\sum_{s=1}^{4} \widehat{Q}^s(\tau)}_{\text{Avg. local quantiles}} \stackrel{??}{=} \underbrace{\widehat{Q}(\tau)}_{\text{global quantile}}$$

$X_i \sim N(0,1)$. $N = 2^{15}$. Repeat 20 times (20 black curves).

$$Q(\tau) = \Phi^{-1}(\tau) \text{ v.s. } m^{-1} \sum_{s=1}^{m} \widehat{Q}^s(\tau)$$







The relative size of m to N matters!

Challenges

- When does the D&C algorithm work uniformly in τ: $m < m^*$. What is m^*?
- Statistical inference for the whole distribution F?
  - More than conditional mean ▸ literature

Challenges

- When does the D&C algorithm work uniformly in τ: $m < m^*$. What is m^*?
- Statistical inference for the whole distribution F?
  - More than conditional mean

literature

Challenges

- When does the D&C algorithm work uniformly in τ: $m < m^*$. What is m^*?
- Statistical inference for the whole distribution F?
  - More than conditional mean literature

Outline

Two-step procedure

Tuning m and K for the oracle rules

Confidence intervals (CIs)

Simulation

Quantile regression as optimization

Linear model: $Q(x;\tau) = x^\top \boldsymbol{\beta}(\tau)$

Koenker and Bassett (1978): Estimate $\boldsymbol{\beta}(\tau)$ by

$$\widehat{\boldsymbol{\beta}}_{or}(\tau) := \arg\min_{\mathbf{b}} \sum_{i=1}^{N} \rho_\tau(Y_i - \mathbf{b}^\top X_i)$$

where $\rho_\tau(u) := \tau u^+ + (1-\tau)u^-$ 'check function'.

- or: oracle, the best we can obtain with sufficient computational resource
- Optimization problem is convex (but non-smooth)

Computational challenges

$$\widehat{Q}_{or}(x_0;\tau) := x_0^\top \widehat{\beta}_{or}(\tau)$$

for any x_0: a fixed vector

- Computing $\widehat{Q}_{or}(x_0;\tau)$ at a fixed τ requires to load all N data in computer memory, which is infeasible when, e.g. $N =$1TB
- Computing $\widehat{Q}_{or}(x_0;\tau)$ for many τ is impossible

Two-step procedure

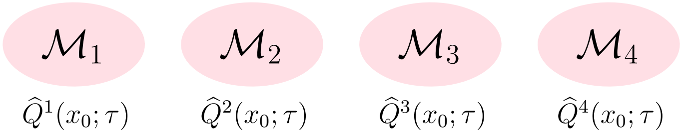
Tuning m and K for the oracle rules

Confidence intervals (CIs)

Simulation

Step 1: D&C algorithm at fixed τ

Local machine $\mathcal{M}_s$ computes $\widehat{Q}^s(x_0;\tau)$, $s = 1, ..., m$ with quantile regression (= solving the optimization problem)



the central computer computes

$$\overline{Q}(x_0;\tau) := \frac{1}{m}\sum_{s=1}^{m}\widehat{Q}^s(x_0;\tau)$$

Tuning m (number of computers)

Greater m

✓ computational efficiency: each local computer processes less data

✗ statistical accuracy: may suffer from great statistical error

$\overline{Q}(x_0; \tau)$ is only for a fixed τ...

Take a grid $\{\tau_1, ..., \tau_K\}$ on $[\tau_L, \tau_U] \subset (0, 1)$, calculate $\{\overline{Q}(x_0; \tau_k)\}_{k=1}^{K}$, then project them to a spline space of τ

Step 2: Quantile projection

B : B-splines defined on G knots in $[\tau_L, \tau_U] \subset (0,1)$

$$\widehat{Q}(x_0;\tau) := \widehat{\boldsymbol{\alpha}}_0^\top \mathbf{B}(\tau)$$

1. Take a grid of quantile levels $\{\tau_1, ..., \tau_K\}$ on $[\tau_L, \tau_U]$, $K > q$
2. Compute $\overline{Q}(x_0;\tau_k)$ for each τ_k (one pass over entire data)
3. (Central machine) Project* $\{\overline{Q}(x_0;\tau_k)\}_k$ on the spline space

$$\widehat{\alpha}_0 := \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \sum_{k=1}^{K} \left(\overline{Q}(x_0;\tau_k) - \boldsymbol{\alpha}^\top \mathbf{B}(\tau_k)\right)^2$$

*with respect to inner product $\langle f, g\rangle_K = \sum_{k=1}^{K} f(\tau_k)g(\tau_k)$

Step 2: Quantile projection

B : B-splines defined on G knots in $[\tau_L, \tau_U] \subset (0,1)$

$$\widehat{Q}(x_0;\tau) := \widehat{\boldsymbol{\alpha}}_0^\top \mathbf{B}(\tau)$$

1. Take a grid of quantile levels $\{\tau_1, ..., \tau_K\}$ on $[\tau_L, \tau_U]$, $K > q$
2. Compute $\overline{Q}(x_0;\tau_k)$ for each τ_k (one pass over entire data)
3. (Central machine) Project* $\{\overline{Q}(x_0;\tau_k)\}_k$ on the spline space

$$\widehat{\alpha}_0 := \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \sum_{k=1}^{K} \left(\overline{Q}(x_0;\tau_k) - \boldsymbol{\alpha}^\top \mathbf{B}(\tau_k)\right)^2$$

*with respect to inner product $\langle f, g\rangle_K = \sum_{k=1}^K f(\tau_k)g(\tau_k)$

Step 2: Quantile projection

B : B-splines defined on G knots in $[\tau_L, \tau_U] \subset (0,1)$

$$\widehat{Q}(x_0;\tau) := \widehat{\boldsymbol{\alpha}}_0^\top \mathbf{B}(\tau)$$

1. Take a grid of quantile levels $\{\tau_1, ..., \tau_K\}$ on $[\tau_L, \tau_U]$, $K > q$
2. Compute $\overline{Q}(x_0;\tau_k)$ for each τ_k (one pass over entire data)
3. (Central machine) Project* $\{\overline{Q}(x_0;\tau_k)\}_k$ on the spline space

$$\widehat{\alpha}_0 := \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \sum_{k=1}^{K} \left(\overline{Q}(x_0;\tau_k) - \boldsymbol{\alpha}^\top \mathbf{B}(\tau_k)\right)^2$$

*with respect to inner product $\langle f, g\rangle_K = \sum_{k=1}^K f(\tau_k)g(\tau_k)$

Step 2: Quantile projection

B : B-splines defined on G knots in $[\tau_L, \tau_U] \subset (0,1)$

$$\widehat{Q}(x_0;\tau) := \widehat{\boldsymbol{\alpha}}_0^\top \mathbf{B}(\tau)$$

1. Take a grid of quantile levels $\{\tau_1, ..., \tau_K\}$ on $[\tau_L, \tau_U]$, $K > q$
2. Compute $\overline{Q}(x_0;\tau_k)$ for each τ_k (one pass over entire data)
3. (Central machine) Project* $\{\overline{Q}(x_0;\tau_k)\}_k$ on the spline space

$$\widehat{\boldsymbol{\alpha}}_0 := \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \sum_{k=1}^{K} \left(\overline{Q}(x_0;\tau_k) - \boldsymbol{\alpha}^\top \mathbf{B}(\tau_k)\right)^2$$

*with respect to inner product $\langle f, g\rangle_K = \sum_{k=1}^{K} f(\tau_k)g(\tau_k)$

Computation of $\widehat{F}_{Y|X}(y|x)$

Given $\widehat{\boldsymbol{\alpha}}_0$, the central computer can compute $\widehat{Q}(x_0;\tau) = \widehat{\boldsymbol{\alpha}}_0^\top \mathbf{B}(\tau)$ for many τ at almost no cost

$$\widehat{F}_{Y|X}(y|x_0) := \tau_L + \int_{\tau_L}^{\tau_U} \mathbf{1}\{\widehat{Q}(x_0;\tau) < y\} d\tau.$$

Tuning K (quantile grid size)

Greater K

✗ computational efficiency: more $\overline{Q}(x_0; \tau_k)$ to compute

✓ statistical accuracy: better projection performance

Two-step procedure

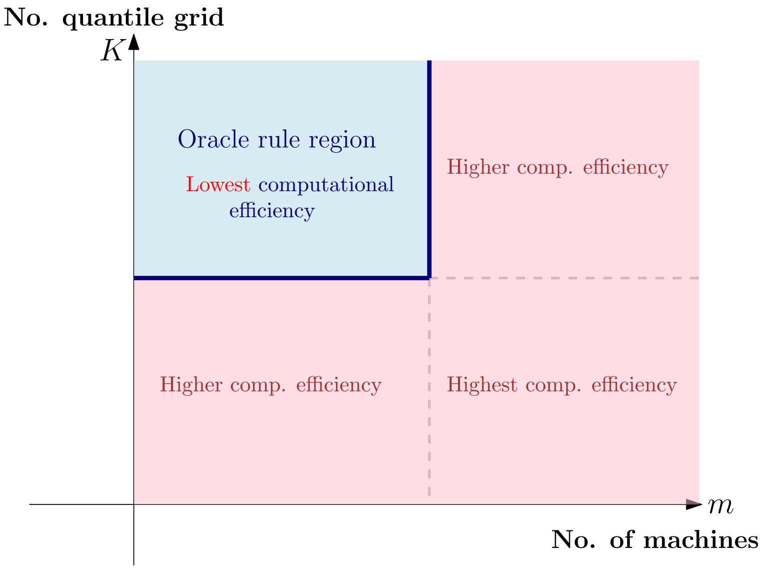
Tuning m and K for the oracle rules

Confidence intervals (CIs)

Simulation

Oracle rule: $\overline{Q}$, $\widehat{Q}$ and $\widehat{F}_{Y|X}$ have the same limiting distribution as the oracles obtained by super computers

How to tune m and K so that the oracle rule holds?


No. quantile grid
K
Oracle rule region
Lowest computational efficiency
Higher comp. efficiency
Higher comp. efficiency
Highest comp. efficiency
m
No. of machines

What are the boundaries?

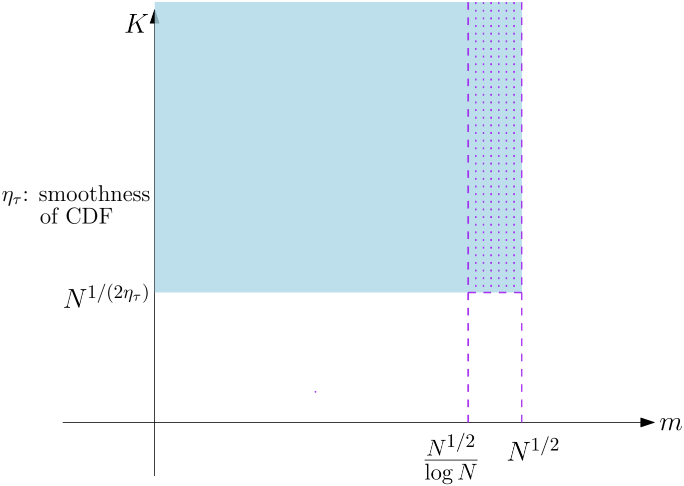


Figure: Blue region: oracle rule holds. Boundary of K is unimprovable.
For m, $\frac{N^{1/2}}{\log N}$ may be improved to $N^{1/2}$, but no further.

Why is $N^{1/2}(\log N)^{-1}$ sufficient?

$$\sqrt{N}(\overline{Q}(x_0;\tau) - Q(x_0;\tau)) = \underbrace{\sqrt{N}\big(\overline{Q}(x_0;\tau) - \mathbb{E}[\overline{Q}(x_0;\tau)]\big)}_{\rightsquigarrow \mathcal{N} \text{ oracle rule}} + \underbrace{\sqrt{N}\,\mathrm{Bias}(\overline{Q}(x_0;\tau))}_{o(1)?}$$

$$\sup_{\tau} \mathrm{Bias}(\overline{Q}(x_0;\tau)) \lesssim \frac{\log n}{n} \ll \frac{1}{\sqrt{N}} \asymp \text{rate of } \mathrm{SD}(\overline{Q}(x_0;\tau))$$

Hence, $m = o(N^{1/2}(\log N)^{-1})$, if we recall $n = N/m$

m cannot go beyond $N^{1/2}$

For some distribution, the bias is bounded from below:

$$\frac{N}{m} = \underbrace{\frac{1}{n} \lesssim \text{Bias of } \overline{Q}(x_0;\tau)}_{\text{computational limit}}$$

If $m \gtrsim \sqrt{N}$, then $\text{Bias}(\overline{Q}(x_0;\tau)) \gtrsim \frac{1}{n} \gtrsim \frac{1}{\sqrt{N}}$,

$\sqrt{N}\,\text{Bias}(\overline{Q}(x_0;\tau))$ is nonvanishing, so the oracle rule fails

m cannot go beyond $N^{1/2}$

For some distribution, the bias is bounded from below:

$$\frac{N}{m} = \underbrace{\frac{1}{n} \lesssim \text{Bias of } \overline{Q}(x_0;\tau)}_{\text{computational limit}}$$

If $m \gtrsim \sqrt{N}$, then $\text{Bias}(\overline{Q}(x_0;\tau)) \gtrsim \frac{1}{n} \gtrsim \frac{1}{\sqrt{N}}$,

$\sqrt{N}\,\text{Bias}(\overline{Q}(x_0;\tau))$ is nonvanishing, so the oracle rule fails

Oracle confidence intervals

Oracle rules: asymptotic normality of $\overline{Q}(x_0;\tau)$ and $\widehat{F}_{Y|X}(y|x_0)$

$$Q(x_0;\tau) : \left[\overline{Q}(x_0;\tau) \pm z_{1-\alpha/2}\, N^{1/2}\mathrm{SD}(\overline{Q}(x_0;\tau))\right]$$

$$F_{Y|X}(y|x_0) : \left[\widehat{F}_{Y|X}(y|x_0) \pm z_{1-\alpha/2}\, N^{1/2}\mathrm{SD}(\widehat{F}_{Y|X}(y|x_0))\right]$$

- $z_{1-\alpha/2}$: critical value from standard normal, $\alpha = 5\%$
- $\mathrm{Var}(\overline{Q}(x_0;\tau))$ and $\mathrm{Var}(\widehat{F}_{Y|X}(y|x_0))$ depending on the underlying distribution are usually unknown

$$\overline{Q}(x_0;\tau) = m^{-1}\sum_{s=1}^{m} \underbrace{\widehat{Q}^s(x_0;\tau)}_{\text{i.i.d. "samples"}} \text{ is an "average"}$$

Central computer has $\widehat{Q}^s(x_0;\tau_k)$ i.i.d. and close to $\mathcal{N}$ for each machine $s = 1, ..., m$ and grid point $k = 1, ..., K$

- $\widehat{\sigma}^2_{0,\tau} = (m-1)^{-1} \sum_{s=1}^{m} \left(\widehat{Q}^s(x_0;\tau) - \overline{Q}(x_0;\tau) \right)^2$
- Small m: the distribution of $\widehat{Q}^s(x_0;\tau)$ is "close" to normal

$$\left[\overline{Q}(x_0;\tau) \pm m^{-1/2} t_{m-1,1-\alpha/2} \, \widehat{\sigma}_{0,\tau} \right] \qquad (t\text{-quantile})$$

- Large m: t_{m-1} is "close" to standard normal

$$\left[\overline{Q}(x_0;\tau) \pm m^{-1/2} z_{1-\alpha/2} \, \widehat{\sigma}_{0,\tau} \right] \qquad (N\text{-quantile})$$

This cannot be extended to the projection estimator, for which we need to deal with all $\tau \in [\tau_L, \tau_U]$

- $\widehat{\sigma}^2_{0,\tau} = (m-1)^{-1} \sum_{s=1}^{m} \left(\widehat{Q}^s(x_0;\tau) - \overline{Q}(x_0;\tau) \right)^2$

- Small m: the distribution of $\widehat{Q}^s(x_0;\tau)$ is "close" to normal

$$\left[\overline{Q}(x_0;\tau) \pm m^{-1/2} t_{m-1,1-\alpha/2} \, \widehat{\sigma}_{0,\tau} \right] \qquad (t\text{-quantile})$$

- Large m: t_{m-1} is "close" to standard normal

$$\left[\overline{Q}(x_0;\tau) \pm m^{-1/2} z_{1-\alpha/2} \, \widehat{\sigma}_{0,\tau} \right] \qquad (N\text{-quantile})$$

This cannot be extended to the projection estimator,
for which we need to deal with all $\tau \in [\tau_L, \tau_U]$

- $\widehat{\sigma}^2_{0,\tau} = (m-1)^{-1} \sum_{s=1}^{m} \left(\widehat{Q}^s(x_0;\tau) - \overline{Q}(x_0;\tau)\right)^2$
- Small m: the distribution of $\widehat{Q}^s(x_0;\tau)$ is "close" to normal

$$\left[\overline{Q}(x_0;\tau) \pm m^{-1/2} t_{m-1,1-\alpha/2}\, \widehat{\sigma}_{0,\tau}\right] \qquad (t\text{-quantile})$$

- Large m: t_{m-1} is "close" to standard normal

$$\left[\overline{Q}(x_0;\tau) \pm m^{-1/2} z_{1-\alpha/2}\, \widehat{\sigma}_{0,\tau}\right] \qquad (N\text{-quantile})$$

This cannot be extended to the projection estimator,
for which we need to deal with all $\tau \in [\tau_L, \tau_U]$

- $\widehat{\sigma}^2_{0,\tau} = (m-1)^{-1} \sum_{s=1}^{m} \left(\widehat{Q}^s(x_0;\tau) - \overline{Q}(x_0;\tau) \right)^2$

- Small m: the distribution of $\widehat{Q}^s(x_0;\tau)$ is "close" to normal

$$\left[\overline{Q}(x_0;\tau) \pm m^{-1/2} t_{m-1,1-\alpha/2}\, \widehat{\sigma}_{0,\tau} \right] \qquad (t\text{-quantile})$$

- Large m: t_{m-1} is "close" to standard normal

$$\left[\overline{Q}(x_0;\tau) \pm m^{-1/2} z_{1-\alpha/2}\, \widehat{\sigma}_{0,\tau} \right] \qquad (N\text{-quantile})$$

This cannot be extended to the projection estimator, for which we need to deal with all $\tau \in [\tau_L, \tau_U]$

Bootstrap

Generate i.i.d. $\{\omega_{s,b}\}_{s=1,\ldots,m,b=1,\ldots,B}$ (independent from data)

$$\overline{Q}^{(b)}(x_0;\tau_k) := \frac{1}{m}\sum_{s=1}^{m}\frac{\omega_{s,b}}{\bar{\omega}_{\cdot,b}}\widehat{Q}^{s}(x_0;\tau_k)$$

$\bar{\omega}_{\cdot,b} = m^{-1}\sum_{s=1}^{m}\omega_{s,b}$

Project $\{\overline{Q}^{(b)}(x_0;\tau_k)\}_{k=1,\ldots,K}$ on spline space (as Step 2)

$$\widehat{Q}^{(b)}(x_0;\cdot) = \widehat{\boldsymbol{\alpha}}_0^{(b)\top}\mathbf{B}(\cdot)$$

$$\widehat{F}_{Y|X}^{(b)}(y|x_0) = \tau_L + \int_{\tau_L}^{\tau_U}\mathbf{1}\{\widehat{Q}^{(b)}(x_0;\tau) < y\}d\tau$$

Bootstrap

Generate i.i.d. $\{\omega_{s,b}\}_{s=1,\dots,m,b=1,\dots,B}$ (independent from data)

$$\overline{Q}^{(b)}(x_0;\tau_k) := \frac{1}{m}\sum_{s=1}^{m}\frac{\omega_{s,b}}{\bar{\omega}_{\cdot,b}}\widehat{Q}^{s}(x_0;\tau_k)$$

$\bar{\omega}_{\cdot,b} = m^{-1}\sum_{s=1}^{m}\omega_{s,b}$

Project $\{\overline{Q}^{(b)}(x_0;\tau_k)\}_{k=1,\dots,K}$ on spline space (as Step 2)

$$\widehat{Q}^{(b)}(x_0;\cdot) = \widehat{\boldsymbol{\alpha}}_0^{(b)\top}\mathbf{B}(\cdot)$$

$$\widehat{F}^{(b)}_{Y|X}(y|x_0) = \tau_L + \int_{\tau_L}^{\tau_U}\mathbf{1}\{\widehat{Q}^{(b)}(x_0;\tau) < y\}d\tau$$

Simulation

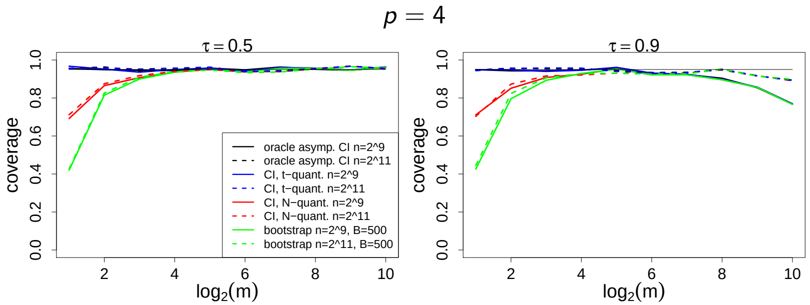
- $Y_i = 0.21 + \boldsymbol{\beta}_{p-1}^{\top} X_i + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, 0.1^2)$
- $X_i \sim \mathcal{U}([0,1]^{p-1})$

▸ Additional information

- $p = 4, 32$
- $Q(x_0; \tau) = 0.21 + \boldsymbol{\beta}_{p-1}^{\top} x_0 + 0.1\Phi^{-1}(\tau)$, Φ: distribution function of $N(0,1)$
- Simulate coverage $= P\{Q(x_0;\tau) \in \mathrm{CI}_\alpha$ for $Q(x_0;\tau)\}$

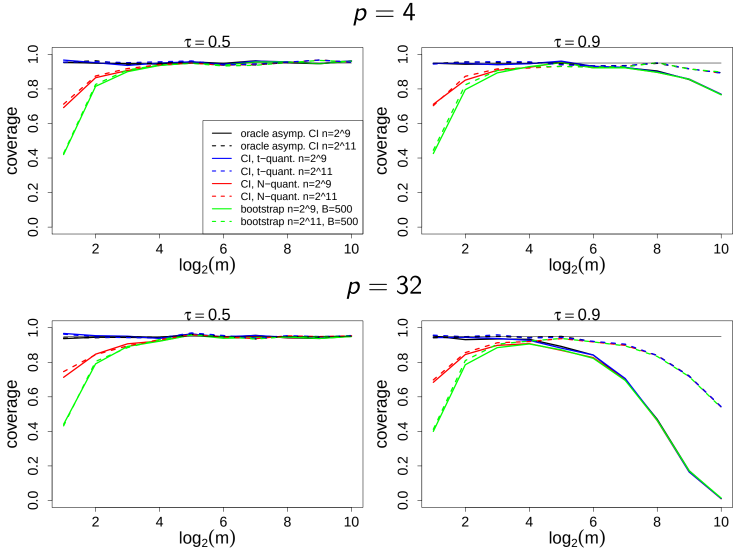
Oracle rule holds if coverage $= 1 - \alpha = 95\%$

CI for $Q(x_0; \tau)$, fixed τ



- $\tau = 0.9$: drop after certain m; $\tau = 0.5$: m not large enough
- m small, bootstrap and $\mathcal{N}$-quant perform badly
- t-quant performs well even for small m
- coverage performs better for bigger n (dashed lines)

CI coverage for $Q(x_0; \tau)$, fixed τ



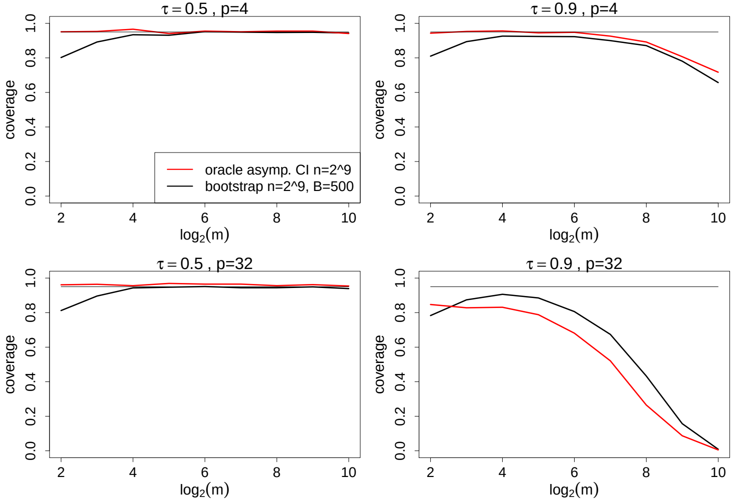
- p increases, coverage drops early for $\tau = 0.9$

Quantile projection

- Quantile grid no. K, no. of knots G are large to undersmooth
  - $K = 65$, $G = 32$ equidistant knots on $[0.05, 0.95]$
  - $\mathbf{B}$: cubic B-spline basis with $\dim(\mathbf{B}) = 28$
- $n = 2^9$
- $y_0 = Q(x_0; \tau)$ so that $F_{Y|X}(y_0|x_0) = \tau$
- Simulate coverage $= P\{\tau \in \mathrm{CI}_\alpha$ for $F_{Y|X}(y_0|x_0)\}$

Oracle rule holds if coverage $= 1 - \alpha = 95\%$

CI coverage for $F_{Y|X}(y_0|x_0) = \tau$



- $p = 4$: coverage drops after certain m
- $p = 32$: n may be too small for the oracle rule to hold; bootstrap CI can be more accurate (correcting for skewness?)

Thank you for your attention!

Volgushev, S., Chao, S.-K. and Cheng, G. (2019). Distributed inference for quantile regression processes. *Annals of Statistics*, 47(3): 1634–1662.

Divide-and-conquer literature

Back

m^* is characterized under different settings, mainly for mean function

- Li et al. ('12): estimate kernel density and distribution parameter, notice that the bias determines $n \gg \sqrt{N} \log\log N$
- Jordan ('13): Bag of Little Bootstraps (e.g. subsample size $n = N^{0.7}$), SVD, denoising problem
- Zhang et al. ('13): empirical risk minimization with parametric smooth loss function, MSE
- Zhang et al. ('15): kernel ridge regression with minimax MSE
- Zhao et al. ('16): PLM, asymp. dist. and minimax MSE
- Shang and Cheng ('17): smoothing spline minimax testing
- Banerjee et al. ('18+): isotonic regression, non-Gaussian limit
- Shi et al. ('18+): M-estimator with cubic rate

This talk: conditional quantile and distribution function, unimprovability for the bound of m, computationally efficient CIs

$\mathcal{P} = \mathcal{P}(\xi_p, M, \bar{f}, \overline{f'}, f_{\min})$: class of distributions of (X, Y) with ▸ Assumption (A) for some constants $0 < \xi_p, M, \bar{f}, \overline{f'} < \infty$ and $f_{\min} > 0$

▸ Back to oracle rule

Assumption (A): data $(X_i, Y_i)_{i=1,\ldots,N}$ are i.i.d. with

▸ Linear models

(A1) Assume that $\|X_i\| \leq \xi_p < \infty$, and that

$$1/M \leq \lambda_{\min}(\mathbb{E}[XX^T]) \leq \lambda_{\max}(\mathbb{E}[XX^T]) \leq M$$

for some fixed constant M.

(A2) The conditional distribution $F_{Y|X}(y|x)$ is twice differentiable w.r.t. y. Denote the corresponding derivatives by $f_{Y|X}(y|x)$ and $f'_{Y|X}(y|x)$. Assume that

$$\bar{f} := \sup_{y,x} |f_{Y|X}(y|x)| < \infty, \quad \overline{f'} := \sup_{y,x} |f'_{Y|X}(y|x)| < \infty$$

uniformly in n.

(A3) Assume that

$$0 < f_{\min} \leq \inf_{\tau \in [\tau_L, \tau_U]} \inf_x f_{Y|X}(Q(x;\tau)|x).$$

$$\sigma_\tau^2(x_0) = x_0^\top J_p(\tau)^{-1}\mathbb{E}[XX^\top]J_p(\tau)^{-1}x_0\ \tau(1-\tau)$$

$$\mathbb{E}\big[\mathbb{G}(\tau)\mathbb{G}(\tau')\big] = x_0^\top J_p(\tau)^{-1}\mathbb{E}\big[XX^\top\big]J_p(\tau')^{-1}x_0\ (\tau\wedge\tau' - \tau\tau')$$

$$\mathbb{E}\big[\mathbb{G}_1(y)\mathbb{G}_1(y')\big] = f_{Y|X}(y|x_0)f_{Y|X}(y'|x_0)\\ \mathbb{E}\big[\mathbb{G}(F_{Y|X}(y|x_0))\mathbb{G}(F_{Y|X}(y'|x_0))\big]$$

where $J_p(\tau) = \mathbb{E}[f_{Y|X}(Q(x_0;\tau)|x_0)XX^\top]$ is the Hessian matrix

▸ Back to oracle rule

Auxiliary information on simulation

- $X_i \sim \mathcal{U}([0,1]^{p-1})$ with covariance $\Sigma_X := \mathbb{E}[X_i X_i^\top]$, where $\Sigma_{jk} = 0.1^2 0.7^{|j-k|}$ for $j, k = 1, ..., p-1$
- $x_0 = (1, (p-1)^{-1/2} \mathbf{l}_{p-1}^\top)^\top$
- $\boldsymbol{\beta}(\tau) = (0.21 + 0.1 \times \Phi_{\sigma=0.1}^{-1}(\tau), \boldsymbol{\beta}_{p-1}^\top)^\top$,

$$\begin{aligned} \boldsymbol{\beta}_3 &= (0.21, -0.89, 0.38)^\top; \\ \boldsymbol{\beta}_{15} &= (\boldsymbol{\beta}_3^\top, 0.63, 0.11, 1.01, -1.79, -1.39, \\ &\qquad 0.52, -1.62, 1.26, -0.72, 0.43, -0.41, -0.02)^\top; \\ \boldsymbol{\beta}_{31} &= (\boldsymbol{\beta}_{15}^\top, 0.21, \boldsymbol{\beta}_{15}^\top)^\top. \end{aligned}$$

Simulation setting

$$\sqrt{N}(\widehat{Q}(x_0;\cdot) - Q(x_0;\cdot))$$

$$= \underbrace{\sqrt{N}\big(\widehat{Q}(x_0;\cdot) - \mathbb{E}[\widehat{Q}(x_0;\cdot)]\big)}_{\rightsquigarrow\ \mathbb{G} \text{ oracle rule}} + \underbrace{\sqrt{N}\,\mathrm{Bias}(\widehat{Q}(x_0;\cdot))}_{\text{force it } o(1)}$$

G: number of knots

$$\sup_\tau \mathrm{Bias}(\widehat{Q}(x_0;\tau)) \leq \text{Bias of projection} + \sup_\tau \mathrm{Bias}(\overline{Q}(x_0;\tau))$$

$$\lesssim G^{-\eta_\tau} + \frac{\log n}{n}$$

$$\ll \frac{1}{\sqrt{N}}$$

this inequality holds when $K \gg G \gg N^{1/(2\eta_\tau)}$ and $m \ll \frac{N^{1/2}}{\log N}$.

Back to oracle rule of $\widehat{Q}$

$$
\begin{aligned}
&\sqrt{N}(\widehat{Q}(x_0;\cdot) - Q(x_0;\cdot)) \\
&= \underbrace{\sqrt{N}\big(\widehat{Q}(x_0;\cdot) - \mathbb{E}[\widehat{Q}(x_0;\cdot)]\big)}_{\rightsquigarrow\ \mathbb{G}\ \text{oracle rule}} + \underbrace{\sqrt{N}\,\mathrm{Bias}(\widehat{Q}(x_0;\cdot))}_{\text{force it } o(1)}
\end{aligned}
$$

G: number of knots

$$
\begin{aligned}
\sup_\tau \mathrm{Bias}(\widehat{Q}(x_0;\tau)) &\le \text{Bias of projection} + \sup_\tau \mathrm{Bias}(\overline{Q}(x_0;\tau)) \\
&\lesssim G^{-\eta_\tau} + \frac{\log n}{n} \\
&\ll \frac{1}{\sqrt{N}}
\end{aligned}
$$

this inequality holds when $K \gg G \gg N^{1/(2\eta_\tau)}$ and $m \ll \frac{N^{1/2}}{\log N}$.

Back to oracle rule of $\widehat{Q}$

- $\text{Bias}(\widehat{Q}(x_0;\cdot)) \gtrsim G^{-\eta_\tau} \gg K^{-\eta_\tau}$ (for a $P \in \mathcal{P}$) for all m

  If $K \lesssim N^{1/(2\eta_\tau)}$, $\sqrt{N}\,\text{Bias}(\widehat{Q}(x_0;\cdot)) \gtrsim \left(\frac{N^{1/(2\eta_\tau)}}{K}\right)^{\eta_\tau}$

- When $K \gg G \gg N^{1/(2\eta_\tau)}$, $\text{Bias}(\widehat{Q}(x_0;\cdot)) \gtrsim \frac{1}{n}$ (for a $P \in \mathcal{P}$)

  If $m \gtrsim N^{1/2}$, $\sqrt{N}\,\text{Bias}(\widehat{Q}(x_0;\cdot)) \gtrsim \frac{\sqrt{N}}{n} \asymp \frac{m}{\sqrt{N}}$

▸ Back to oracle rule for $\widehat{Q}$

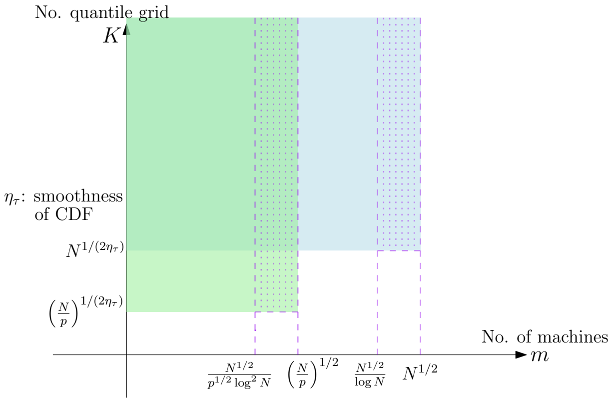
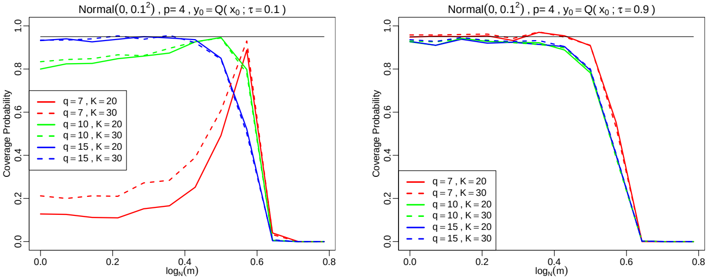


Figure: Oracle rule of linear and nonparametric model.

Back

$F_{Y|X}(y|x)$, $\varepsilon \sim \mathcal{N}(0, 0.1^2)$, **Oracle CI**, $N = 2^{14}$

$p = 4$

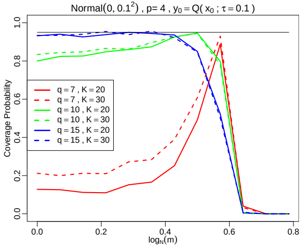


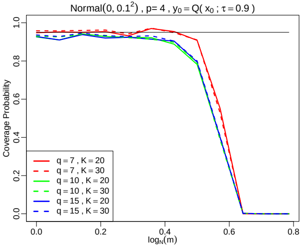
$q = \dim(\mathbf{B})$, $\mathbf{B}$: cubic B-spline basis for projection, K : # quantile grid points

- $\tau = 0.1$: interplay between bias from high m (# machines) and bias from low $q = \dim(\mathbf{B})$ (oversmoothing)
- Either large m or small q corrupts the oracle rule
- Coverage is no longer symmetric in τ

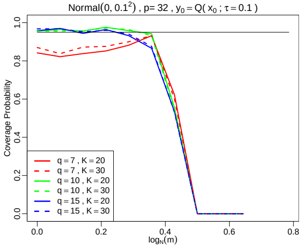
$F_{Y|X}(y|x)$, $\varepsilon \sim \mathcal{N}(0, 0.1^2)$, **Oracle CI,** $N = 2^{14}$

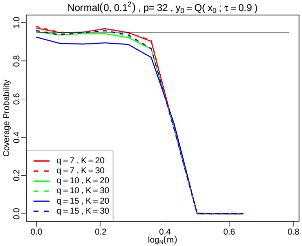
$p = 4$





$p = 32$



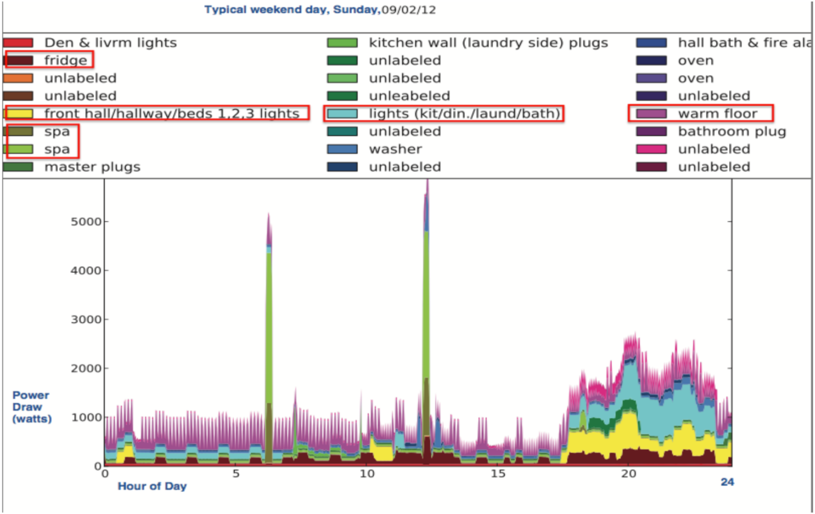


Increase in p lowers both m^* (bad) and q^* (good)

Reference Energy Disaggregation Data Set (REDD)

- Public accessible
- Disaggregated: 30 households, measurements from 24 devise-specific electricity consumption sources: microwave, refrigerator, dishwasher, electronics, lighting...
- Numeric data (Watts), entire data size > 1 TB
- Preliminary idea: compare the distribution of energy consumption across different devices and different time in a day

Back to future study



[Kolter and Johnson, 2011]

▸ Back to future study