Distributed Inference for Quantile Regression Processes

Stanislav Volgushev[1], Shih-Kang Chao[2](skchao74@purdue.edu), Guang Cheng[2]
[1]Department of Statistical Sciences, University of Toronto; [2]Department of Statistics, Purdue University

Objective

Perform inference using distributed quantile regression and its projected process. Particularly, we need *sharp* conditions in (S, K) such that the "oracle rule" holds: $\overline{\beta}(\tau)$ in (7) satisfies (3), and $\widehat{\beta}(\tau)$ in (8) satisfies (4).

Quantile regression

Let $\{(X_i, Y_i)\}_{i=1}^N$ be independent and identical samples in $\mathbb{R}^{d+1}$, where N may be so large that a standalone machine cannot process all the data. Take $\mathcal{T} = [\tau_L, \tau_U]$ with $0 < \tau_L < \tau_U < 1$, estimate for any fixed $\tau \in \mathcal{T}$, the τ-quantile $Q(x;\tau)$ of Y given X:

$$P(Y \le Q(x;\tau) | X = x) = \tau. \quad (1)$$

Koenker and Bassett (1978): if $Q(x;\tau) = \beta(\tau)^\top x$, estimate by

$$\widehat{\beta}_{or}(\tau) := \arg\min_{\mathbf{b}\in\mathbb{R}^m} \sum_{i=1}^N \rho_\tau\{Y_i - \mathbf{b}^\top \mathbf{Z}(X_i)\} \quad (2)$$

where $\rho_\tau(u) := \tau u^+ + (1-\tau)u^-$ 'check function'. $\mathbf{Z}(x) \in \mathbb{R}^m$ are transformations of x, e.g. linear model with fixed/increasing dimension, B-splines, polynomials, trigonometric polynomials

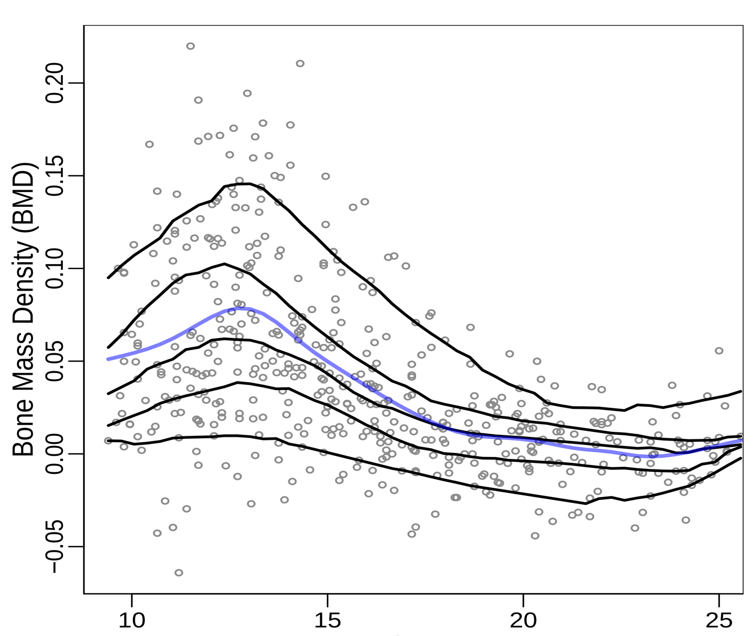


Figure 1: Quantile curves $\widehat{Q}(x;\tau)$ (black, $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$) and the mean curve (blue).

Asymptotics of $\widehat{\beta}_{or}(\tau)$

Under regularity conditions on $\mathbf{Z}$ and the conditional density $f_{Y|X}(y|x)$, for any $x_0 \in \mathcal{X}$ and $\tau \in \mathcal{T}$, $\widehat{\beta}_{or}(\tau)$ has the weak limit (Chao et al., 2016):

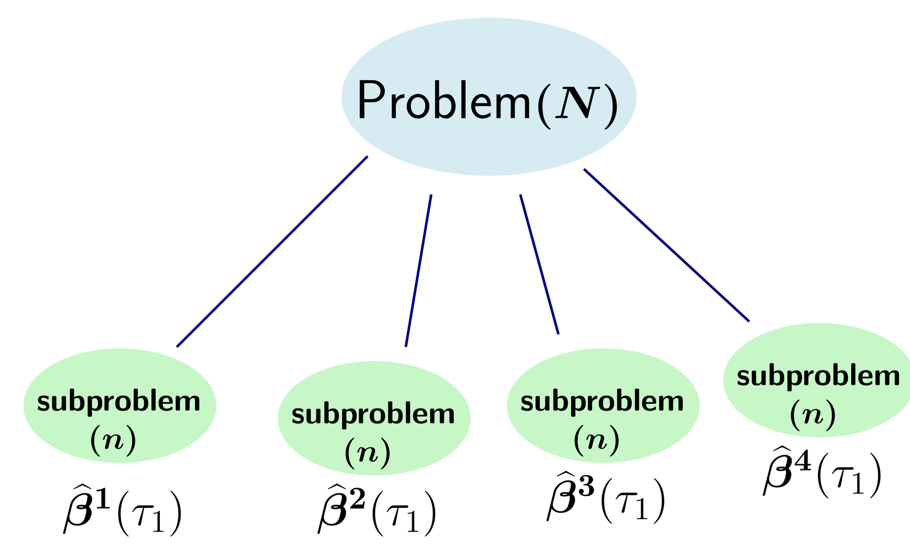
$$\sigma_{m,\tau}^{-1}(x_0)\sqrt{N}\big(\mathbf{Z}(x_0)^\top \widehat{\beta}_{or}(\tau) - Q(x_0;\tau)\big) \rightsquigarrow \mathcal{N}(0,1) \quad (3)$$

$$\sqrt{N}\big(\mathbf{Z}(x_0)^\top \widehat{\beta}_{or}(\cdot) - Q(x_0;\cdot)\big) \rightsquigarrow \mathbb{G}(\cdot) \quad (4)$$

$$\sqrt{N}\big(\widehat{F}^{or}_{Y|X}(\cdot|x_0) - F_{Y|X}(\cdot|x_0)\big) \rightsquigarrow - f_{Y|X}(\cdot|x_0)\mathbb{G}\big(F_{Y|X}(\cdot|x_0)\big), \quad (5)$$

where $\mathbb{G}$ is a centered Gaussian process in $\ell^\infty(\mathcal{T})$ with continuous sample path, $\sigma^2_{m,\tau}(x_0) = \tau(1-\tau)\mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\mathbb{E}[\mathbf{Z}(X)\mathbf{Z}(X)^\top]J_m(\tau)^{-1}\mathbf{Z}(x_0)$.

Quantile D&C and projection



Dividing N samples into S sub-samples.

$$\widehat{\beta}^s(\tau) := \arg\min_{\mathbf{b}\in\mathbb{R}^m} \sum_{i=1}^n \rho_\tau\{Y_{is} - \mathbf{b}^\top \mathbf{Z}(X_{is})\} \quad (6)$$

$$\overline{\beta}(\tau) := \frac{1}{S}\sum_{s=1}^S \widehat{\beta}^s(\tau). \quad (7)$$

However, this is only for a fixed τ! Using projection to avoid repetitively applying D&C. Take $\mathbf{B} := (B_1, ..., B_q)^\top$ B-spline basis defined on equidistant knots $\{t_1, ..., t_G\} \subset \mathcal{T}$ with degree $r_\tau \in \mathbb{N}$,

$$\widehat{\beta}(\tau) := \widehat{\Xi}^\top \mathbf{B}(\tau). \quad (8)$$

Computation of $\widehat{\Xi}$:

(a) Define a grid of quantile levels $\{\tau_1, ..., \tau_K\}$ on $[\tau_L, \tau_U]$, $K > q$. For each τ_k, compute $\overline{\beta}(\tau_k)$ as (7)

(b) Compute for each $j = 1, ..., m$

$$\widehat{\alpha}_j := \arg\min_{\alpha\in\mathbb{R}^q} \sum_{k=1}^K \big(\overline{\beta}_j(\tau_k) - \alpha^\top \mathbf{B}(\tau_k)\big)^2. \quad (9)$$

(c) Set the matrix $\widehat{\Xi} := [\widehat{\alpha}_1\, \widehat{\alpha}_2 \ldots \widehat{\alpha}_m]$.

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

Let $\widehat{\beta}_{or}(\tau)$ and $\widehat{\beta}(\tau)$ be defined in (2) and (8).

$$\widehat{F}^{or}_{Y|X}(y|x_0) := \tau_L + \int_{\tau_L}^{\tau_U} \mathbf{1}\{\mathbf{Z}(x_0)^\top \widehat{\beta}_{or}(\tau) < y\} d\tau. \quad (10)$$

$$\widehat{F}_{Y|X}(y|x_0) := \tau_L + \int_{\tau_L}^{\tau_U} \mathbf{1}\{\mathbf{Z}(x_0)^\top \widehat{\beta}(\tau) < y\} d\tau. \quad (11)$$

where $0 < \tau_L < \tau_U < 1$.

Oracle rule region

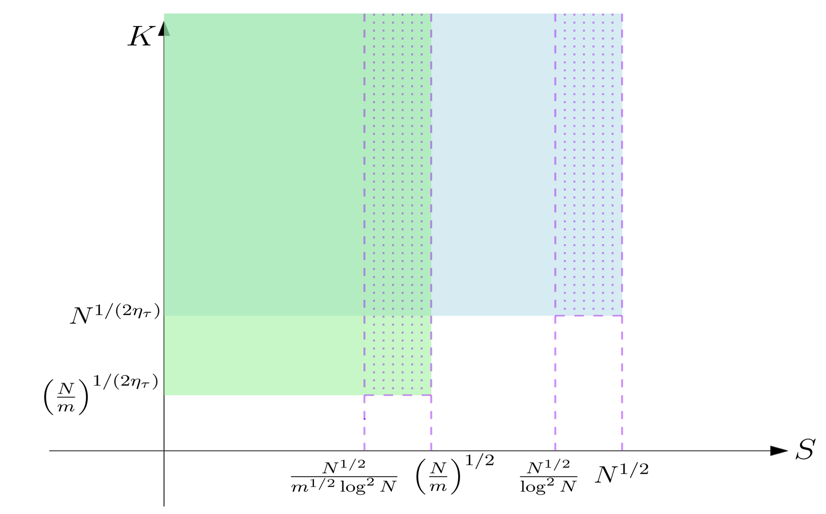


Figure 2: Necessary and sufficient conditions on (S, K) for the oracle rule ($\overline{\beta}(\tau)$ in (7) satisfies (3), and $\widehat{\beta}(\tau)$ in (8) satisfies (4)) in linear models with fixed dimension $m < \infty$ (Blue) and B-spline nonparametric models $m \to \infty$ (Green).
The dotted region is the discrepancy between the sufficient and necessary conditions.

Simulated coverage probabilities of confidence interval based on $\overline{\beta}(\tau)$

We generate data from $Y_i = 0.21 + \beta_{m-1}^\top X_i + \varepsilon_i$, for $m = 4, 16, 32$. $X_i \sim \mathcal{U}([0,1]^{m-1})$ with covariance matrix $\Sigma_X := \mathbb{E}[X_i X_i^\top]$, $\Sigma_{jk} = 0.1^2 0.7^{|j-k|}$ for $j, k = 1, ..., m-1$. The error $\varepsilon \sim \mathcal{N}$ or $\varepsilon \sim$ EXP (skewed). $x_0 = (1, (m-1)^{-1/2}\mathbf{1}^\top_{m-1})^\top$. The 95% coverage probability of the confidence interval from (3) using $\overline{\beta}(\tau)$:

$$P\Big\{x_0^\top \beta(\tau) \in \Big[x_0^\top \overline{\beta}(\tau) \pm N^{-1/2} f^{-1}_{\varepsilon,\tau}\sqrt{\tau(1-\tau)x_0^\top \Sigma_X^{-1} x_0}\,\Phi^{-1}(0.975)\Big]\Big\}$$

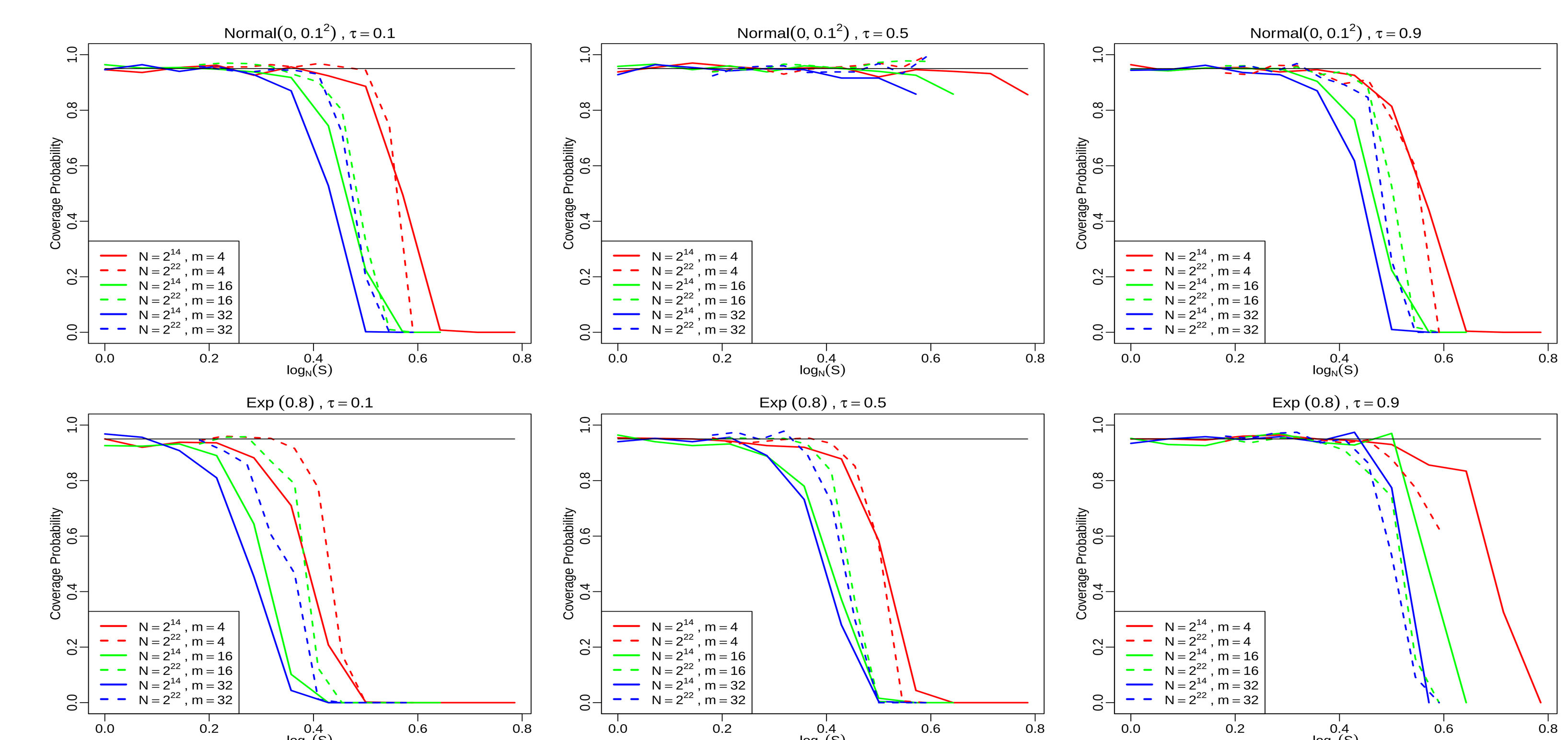


Figure 3: Phase transition: coverage probability drop to 0 after certain threshold S^*. When model dimension m increases, S^* decreases. As N increases, S^* gets closer to $N^{1/2}$ (cf. blue region in Figure 2). In the normal case, the coverage is symmetric in τ.

Simulated coverage probabilities of confidence interval based on $\widehat{F}_{Y|X}(y|x)$

Same setting for (X_i, Y_i) as above. Take $\mathbf{B}$: cubic B-spline with $q = \dim(\mathbf{B})$ defined on $G = 4 + q$ equidistant knots on $[\tau_L, \tau_U]$. We require $K > q$ so that $\widehat{\beta}(\tau)$ is computable. $N = 2^{14}$. $y_0 = Q(x_0;\tau)$ so that $F_{Y|X}(y_0|x_0) = \tau$. The 95% coverage probability of the confidence interval from (5) using $\widehat{F}_{Y|X}(y|x)$ is

$$P\Big\{\tau \in \Big[\widehat{F}_{Y|X}(Q(x_0;\tau)|x_0) \pm N^{-1/2}\sqrt{\tau(1-\tau)x_0^\top \Sigma_X^{-1} x_0}\,\Phi^{-1}(0.975)\Big]\Big\}$$

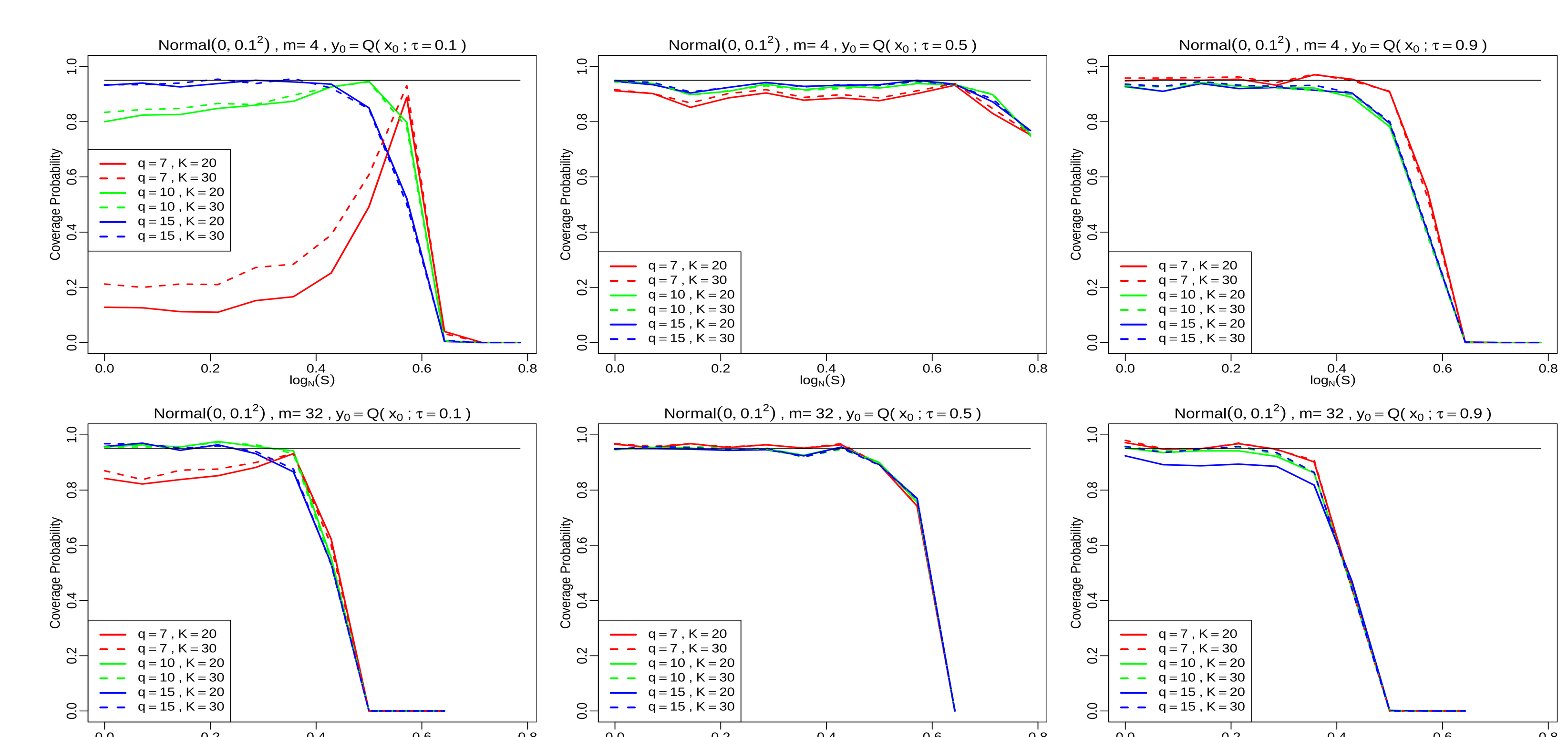


Figure 4: Phase transition: coverage probability drop to 0 after either thresholds S^* or q^*. Increase in model dimension m lowers both S^* and q^*. Increase in q and K improves the coverage probability. Projection induces additional error causing the normal case asymmetric in τ.