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: \(\hat{\beta}(\tau)\) in (7) satisfies (3), and \(\hat{\beta}(\tau)\) 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 stand-alone machine cannot process all the data. Take \(T = (\tau_1, \tau_2)\) with \(0 < \tau_1 < \tau_2 < 1\), estimate for any fixed \(\tau \in T\), the \(\tau\)-quantile \(Q(X; \tau)\) of \(Y\) given \(X\):

\[
P(Y \leq Q(x; \tau)|X = x) = \tau.
\]

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

\[
\hat{\beta}_m(\tau) := \arg \min_{\beta \in \mathbb{R}^d} \frac{1}{N} \sum_{i=1}^N \rho_\tau(Y_i - \beta^T Z(X_i))
\]

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

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

Asymptotics of \(\hat{\beta}_m(\tau)\)

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

\[
\sigma_{\alpha}^{-1}(x_0) \sqrt{N} \left(\hat{\beta}_m(\tau) - Q(X; \tau)\right) \xrightarrow{d} N(0, 1)
\]

\[
\sqrt{N} \left(\hat{\beta}_m(\tau) - Q(X; \tau)\right) \xrightarrow{d} G(\cdot)
\]

\[
\sqrt{N} \left(\hat{\beta}_m(\tau) - Q(X; \tau)\right) \xrightarrow{d} - f_{Y|X}(x_0) \mathbb{G}(F_{Y|X}(x_0))
\]

where \(\mathbb{G}\) is a centered Gaussian process in \(\mathbb{R}^m(T)\) with continuous sample path, \(\sigma_{\alpha}^{-1}(x_0) = \tau(1 - \tau)Z(x_0)^T D_m(\tau)^{-1}E[Z|X]Z(X)^T Zm(\tau)^{-1}Z(x_0)\).

Quantile D&C and projection

Dividing \(N\) samples into \(S\) sub-samples.

\[
\hat{\beta}^*(\tau) := \arg \min_{\beta \in \mathbb{R}^d} \frac{1}{S} \sum_{j=1}^S \rho_\tau\left(\left\{Y_{j,n} - \beta^T Z(X_{j,n})\right\}\right)
\]

\[
\hat{\beta}(\tau) := \frac{1}{S} \sum_{i=1}^S \hat{\beta}^*(\tau).
\]

However, this is only for a fixed \(\tau\)! Using projection to avoid repetitively applying D&C. Take \(B := (B_1, ..., B_3)\) B-spline basis defined on equidistant knots \(\{t_1, ..., t_3\} \subset T\) with degree \(r\), in \(H^r\),

\[
\hat{\beta}(\tau) := \Xi^T B(\tau).
\]

Computation of \(\hat{\Xi}\):

(a) Define a grid of quantile levels \(\{\tau_1, ..., \tau_K\}\) on \([\tau_1, \tau_2]\), \(K > q\). For each \(\tau_q\), compute \(\hat{\beta}(\tau_q)\) as (7)

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

\[
\hat{\alpha}_j := \arg \min_{\alpha \in \mathbb{R}^d} \sum_{k=1}^K (\beta_k(\tau_k) - \alpha^T B(\tau_k))^2.
\]

(c) Set the matrix \(\hat{\Xi} := [\hat{\alpha}_1 \hat{\alpha}_2 ... \hat{\alpha}_m].\)

Computation of \(\hat{\beta}_{Y|X}(\tau|x)\)

Let \(\hat{\beta}_m(\tau)\) and \(\hat{\beta}(\tau)\) be defined in (2) and (8).

\[
\hat{\beta}_{Y|X}(\tau|x) := \tau L + \int_{\tau L}^{\tau U} 1\{Z(x) > \beta_\tau(\tau)\} Y d\tau.
\]

\[
\hat{\beta}_{Y|X}(\tau|x) := \tau L + \int_{\tau L}^{\tau U} 1\{Z(x) > \beta(\tau)\} Y d\tau.
\]

where \(0 < \tau_L < \tau_U < 1\).

Oracle rule region

We generate data from \(Y_j = 0.2 + \beta_{Y|X}(x_j) + \varepsilon_j\) for \(m = 4, 16, 32, \tau_0 = 0.7^{j-1}\) with covariance matrix \(\Sigma_X := \mathbb{E}[X^T X]\), \(\Sigma_{\varepsilon} = 0.1^2 0.7^{j-1}\) for \(j, k = 1, ..., m - 1\). For \(R = \mathbb{E} \{0.05\} \varepsilon \sim \exp(0.05)\). \(x_0 = 1, (m - 1)^{-1/4}\). The 95% coverage probability of the confidence interval from (3) using \(\hat{\beta}(\tau)\):

\[
P\{x_0 Q(x_0; \tau) \in \left[\hat{x}_0(\hat{\beta}(\tau)) \pm N^{-1/2} f_{\beta}(1 - \tau) x_0 \Sigma_{x_0} x_0^{-1}(0.975)\right]\}
\]

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

Same setting as \((X_i, Y_i)\) as above. Take \(B\) cubic B-spline with \(q = \dim(B)\) defined on \(G = 4 + q\). Assume quantile knots on \([\tau_1, \tau_2]\). We require \(K > q\) so that \(\hat{\beta}(\tau)\) is computable. \(N = 2^{11}\). \(y_0 = Q(x_0; \tau)\) so that \(f_{Y|X}(x_0; \tau) = \tau\).

The 95% coverage probability of the confidence interval from (5) using \(\hat{F}_{Y|X}(y|x)\) is

\[
P\{\tau \in \left[\hat{F}_{Y|X}(Q(x_0; \tau)|x_0) \pm N^{-1/2} \sqrt{\tau(1 - \tau) x_0 \Sigma_{x_0} x_0^{-1}(0.975)}\right]\}
\]

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

Figure 2: Necessary and sufficient conditions on \((S, K)\) for the oracle rule \(\hat{\beta}(\tau)\) (in (7) satisfies (3), and \(\hat{\beta}(\tau)\) 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.

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 \(\tau\).

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 \(\tau\).