A Nonparametric Survival Function Estimator via Censored Kernel Quantile Regressions

Korea University, University of Arizona, and North Carolina State University

Supplementary Material

This document contains proofs of Theorem 1 and 2, and the detailed steps for the two dimensional solution surface algorithm of the censored kernel quantile regression.

S1 Proof of Theorem 1

The estimated regression function $\hat{f}(\mathbf{x})$ whose form of (6) can be rewritten as

$$\begin{aligned}\hat{f}(\mathbf{x}) &= \left[\hat{f}(\mathbf{x}) - \frac{\lambda^\ell}{\lambda}\hat{f}^\ell(\mathbf{x})\right] + \frac{\lambda^\ell}{\lambda}\hat{f}^\ell(\mathbf{x}) \\ &= \frac{1}{\lambda}\left[(\hat{\theta}_0 - \hat{\theta}_0^\ell) + \sum_{j\in\mathcal{E}^\ell}(\hat{\theta}_j - \hat{\theta}_j^\ell)K(\mathbf{x},\mathbf{x}_j) + (\tau - \tau_\ell)\sum_{j\notin\mathcal{E}^\ell} w_j K(\mathbf{x},\mathbf{x}_j)\right] + \frac{\lambda^\ell}{\lambda}\hat{f}^\ell(\mathbf{x}). \quad \text{(M.1)}\end{aligned}$$

For $i \in \mathcal{E}^\ell$, we have $y_i - \hat{f}^\ell(\mathbf{x}_i) = y_i - \hat{f}(\mathbf{x}_i) = 0$ which leads (M.1) to:

$$(\lambda - \lambda^\ell)y_i - (\tau - \tau^\ell)\sum_{j\notin\mathcal{E}^\ell}\omega_i K(\mathbf{x}_i,\mathbf{x}_j) = \hat{\theta}_0 - \hat{\theta}_0^\ell + \sum_{j\in\mathcal{E}^\ell}(\hat{\theta}_j - \hat{\theta}_j^\ell)K(\mathbf{x}_i,\mathbf{x}_j), \quad \forall i \in \mathcal{E}^\ell. \quad \text{(M.2)}$$

Moreover, the solution must satisfies $\sum_{i=1}^n \hat{\theta}_i = \sum_{i=1}^n \hat{\theta}_i^\ell = 0$ by Krush-Kuhn-Tucker conditions and hence we have

$$-(\tau - \tau^\ell)\sum_{j\notin\mathcal{E}^\ell} w_i = \sum_{j\in\mathcal{E}^\ell}(\hat{\theta}_j - \hat{\theta}_j^\ell). \quad \text{(M.3)}$$

Together (M.2) and (M.3) form a set of $(|\mathcal{E}^\ell| + 1)$ linear equations, which can be expressed in a matrix form as $\mathbf{B}_\ell(\hat{\boldsymbol{\theta}}_{0,\mathcal{E}} - \hat{\boldsymbol{\theta}}_{0,\mathcal{E}}^\ell) = \mathbf{A}_\ell\boldsymbol{\Delta}$. Finally, the linear update equation (8) follows.

S2 Proof of Theorem 2

We start with introducing some notations:

$$R_{\text{reg}}(f;\tau) = E\left[\frac{\delta}{G(Y)}\rho_\tau(Y - f(\mathbf{x}))\right] + \frac{\alpha_n}{2}\|f\|^2_{\mathcal{H}_K},$$
$$R_{n,\text{reg}}(f;\tau) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{G(Y_i)}\rho_\tau(Y_i - f(\mathbf{x}_i)) + \frac{\alpha_n}{2}\|f\|^2_{\mathcal{H}_K}.$$

For $\epsilon > 0$ fixed, we have a positive integer N_1 such that $\alpha_n \sup\|f\|^2_{\mathcal{H}_K} < \frac{\epsilon}{2}, \forall n > N_1$ by (A1). In addition, we have another N_2 by (A2) and (A3) such that $|\hat{R}_{n,\text{reg}}(f;\tau) - R_{n,\text{reg}}(f;\tau)| < \epsilon/8, \forall n > N_2$. For $\forall n > N = \max(N_1, N_2)$,

$$\begin{aligned}
\sup_\tau \left|R^*(\hat{f}_\tau;\tau) - R^*(f^*_\tau;\tau)\right| =& \sup_\tau \left|R(\hat{f}_\tau;\tau) - R(f^*_\tau;\tau)\right| \\
\le& \sup_\tau \left|R(\hat{f}_\tau;\tau) - \hat{R}_{n,\text{reg}}(\hat{f}_\tau;\tau) + \hat{R}_{n,\text{reg}}(f^*_\tau;\tau) - R(f^*_\tau;\tau)\right| \\
\le& \sup_\tau \left|R_{\text{reg}}(\hat{f}_\tau;\tau) - \hat{R}_{n,\text{reg}}(\hat{f}_\tau;\tau) + \hat{R}_{n,\text{reg}}(f^*_\tau;\tau) - R_{\text{reg}}(f^*_\tau;\tau)\right| + \frac{\epsilon}{4} \\
\le& \sup_\tau \left|R_{\text{reg}}(\hat{f}_\tau;\tau) - R_{n,\text{reg}}(\hat{f}_\tau;\tau) + R_{n,\text{reg}}(f^*_\tau;\tau) - R_{\text{reg}}(f^*_\tau;\tau)\right| + \frac{\epsilon}{2} \\
\le& 2\sup_{f,\tau} \left|R_n(f;\tau) - R(f;\tau)\right| + \frac{\epsilon}{2}.
\end{aligned}$$

Notice that there exists a constant K such that $\forall\varphi \in \mathscr{L}$ is bounded by K under the assumptions 1 and 2. Now applying Theorem 24 of Pollard (1984)

$$\begin{aligned}
P\left(\sup_\tau \left|R^*(\hat{f}_\tau;\tau) - R^*(f^*_\tau;\tau)\right| > \epsilon\right) \le& P\left(2\sup_{f,\tau}\left|R_n(f;\tau) - R(f;\tau)\right| + \frac{\epsilon}{2} > \epsilon\right) \\
\le& P\left(\sup_{f,\tau}\left|R_n(f;\tau) - R(f;\tau)\right| > \frac{\epsilon}{4}\right) \\
\le& 8N_\infty(\tfrac{\epsilon}{32}, n, \mathscr{L})\exp\left(-\frac{n\epsilon^2}{2^{11}K^2}\right),
\end{aligned}$$

where $N_\infty(\epsilon, n, \mathscr{L}) := \sup_{\mathbb{P}_n} N_\infty(\epsilon, \mathbb{P}_n, \mathscr{L})$ with the empirical measure $\mathbb{P}_n$ denotes uniform ℓ_∞-covering number. In order to obtain the ℓ_∞-covering number bound, we firstly show that $\varphi(\cdot)$ satisfies the *Lipschitz condition*.

$$\varphi(\mathbf{Z}; f_1, \tau) - \varphi(\mathbf{Z}; f_2, \tau) = \frac{\delta}{G(Y|\mathbf{X})}\left|\rho_\tau(Y, f_1) - \rho_\tau(Y, f_2)\right| \le \frac{\delta}{G(Y|\mathbf{X})}\left|f_1 - f_2\right|,$$

and $E[\delta/G(Y|\mathbf{X})] = 1$. As pointed out by Zhang (2002), since the *Lipschitz condition* holds for $\varphi(\cdot) \in \mathscr{L}$, the uniform ℓ_∞-covering number bound of $\mathscr{L}$ can be obtained by using the one of a class of function f, denoted by $\mathscr{F}$. Moreover, the two have the same growth rate in terms of the sample size n. By (A1), we have $\mathscr{F} = \{f : \|f\|_{\mathcal{H}_K} \sup_x \|K(\cdot, \mathbf{x})\|_{\mathcal{H}_K} \leq M_1\}$ for a constant M_1. By the theorem 4 of Zhang (2002) we have the following.

$$\log N_\infty(\epsilon, n, \mathscr{F}) \leq M_2 M_1^2 \frac{\log(2 + \frac{M_1}{\epsilon}) + \log n}{n\epsilon}$$

where M_1 is a constant. Therefore $\log N_\infty(\epsilon, n, \mathscr{L}) = O(\log n/n)$, which completes the proof by Borel-Cantelli Lemma.

S3 Algorithm for Two-Dimensional Solution Surface

Due to the joint piecewise linearity (8), we can build the entire solution surface on $\mathcal{S}^\ell$ and therefore the main step of the two-dimensional solution surface algorithm is to obtain a set $\mathcal{S}^\ell$ explicitly. We developed the proposed algorithm in `R` language and is available from the authors upon request.

S3.1 Initialization

For simplicity, we assume that the data are properly ordered as $y_1 > y_2 > \cdots > y_n$. The initialization step sets the starting values of λ^1 and τ^1 with the associated estimates and sets, denoted by $\hat{\boldsymbol{\theta}}^1 = (\hat{\theta}_0^1, \hat{\theta}_1^1, \cdots, \hat{\theta}_n^1)^T$ and $\mathcal{E}^1, \mathcal{L}^1, \mathcal{R}^1$, respectively. For $k = 1, \cdots, n$, we first compute $\mathcal{L}_k^1 = \{1, \cdots, k\}$, $\mathcal{R}_k^1 = \{k+1, \cdots, n\}$, and

$$\lambda_k^1 = \max_{i \in \mathcal{L}_k^1, j \in \mathcal{R}_k^1} \frac{q_k(\mathbf{x}_i) - q_k(\mathbf{x}_j)}{y_i - y_j}$$

where $q_k(\mathbf{x}) = -(1-\tau_k^1) \sum_{i \in \mathcal{L}_k^1} \omega_i K(\mathbf{x}, \mathbf{x}_i) + \tau_k^1 \sum_{j \in \mathcal{R}_k^1} \omega_j K(\mathbf{x}, \mathbf{x}_j)$ with $\tau_k^1 = \sum_{i \in \mathcal{L}_k^1} \omega_i / \sum_{i=1}^n \omega_i$. Now, define $k^* = \text{argmax}_{k \in \{1, \cdots, n-1\}} \lambda_k^1$ then we have indices i^* and j^* such that $\lambda_k^* = \frac{q_k(\mathbf{x}_{i^*}) - q_k(\mathbf{x}_{j^*})}{y_{i^*} - y_{j^*}}$. Finally, the initial value of λ and τ are given by

$$\lambda^1 = \lambda_{k^*}^1 \quad \text{and} \quad \tau^1 = \tau_{k^*}^1,$$

with the associated estimates as

$$\hat{\theta}_i^1 = \begin{cases} -(1-\tau^1)\omega_i & \text{if } i \in \mathcal{L}_{k^*}^1 \\ \tau^1 \omega_i & \text{if } i \in \mathcal{R}_{k^*}^1 \end{cases} \quad \text{and} \quad \hat{\theta}_0^1 = y_{i^*} - q_{k^*}(\mathbf{x}_{i^*}) = y_{j^*} - g_{k^*}(\mathbf{x}_{j^*}),$$

and sets as

$$\mathcal{E}^1 = \{i^*, j^*\},\ \mathcal{L}^1 = \mathcal{L}^1_{k^*} \setminus i^*, \quad \text{and} \quad \mathcal{R}^1 = \mathcal{R}^1_{k^*} \setminus j^*.$$

We point out that any solution for $\lambda > \lambda^1$ (regardless of the value of τ) is trivial in the sense that its elbow set is empty and the solution is readily determined from the definition of left and right set. Defining $\mathcal{Q}$ to denote the region on the $(\lambda \times \tau)$-plane having meaningful solutions as $\mathcal{Q} = \{(\lambda, \tau) : 0 \le \lambda \le \lambda^1, 0 \le \tau \le \tau_0\}$, it is enough for the proposed algorithm to search the CKQR solutions only on $\mathcal{Q}$.

S3.2 Updating $\mathcal{S}^\ell$

In Section 3.1, the $\mathcal{S}^\ell$ is defined as a subregion on the $(\lambda \times \tau)$-plane such that all the sets remain the same as $\mathcal{E}^\ell, \mathcal{L}^\ell$, and $\mathcal{R}^\ell$. Therefore we have the following constraints in order to define the $\mathcal{S}^\ell$.

First, *Event 1* can not be occurred as long as θ_i fails to reach its lower bound $-\omega_i(1-\tau)$ for all $i \in \mathcal{E}^\ell$. That is, by (8)

$$\hat{\theta}_i^\ell + g_{i1}^\ell(\lambda - \lambda^\ell) + g_{i2}^\ell(\tau - \tau^\ell) \ge -\omega_i(1-\tau), \qquad \forall i \in \mathcal{E}^\ell. \tag{M.4}$$

We also have similar constraints in order to prevent *Event 2* from happening as follows.

$$\hat{\theta}_i^\ell + g_{i1}^\ell(\lambda - \lambda^\ell) + g_{i2}^\ell(\tau - \tau^\ell) \le \omega_i\tau, \qquad \forall i \in \mathcal{E}^\ell. \tag{M.5}$$

For preventing the last *Event 3*, $f^\ell(\mathbf{x}_i) > y_i$ for $i \in \mathcal{L}^\ell$ and $f^\ell(\mathbf{x}_j) < y_j$ for $j \in \mathcal{R}^\ell$. Employing (11) we have

$$\{y_i - h_1^\ell(\mathbf{x}_i)\}\lambda - h_2^\ell(\mathbf{x}_i)\tau \le \lambda^\ell\{f^\ell(\mathbf{x}_i) - h_2^\ell(\mathbf{x}_i)\} - \tau^\ell h_2^\ell(\mathbf{x}_i), \quad \forall i \in \mathcal{L}^\ell, \tag{M.6}$$

$$\{y_j - h_1^\ell(\mathbf{x}_j)\}\lambda - h_2^\ell(\mathbf{x}_j)\tau \ge \lambda^\ell\{f^\ell(\mathbf{x}_j) - h_2^\ell(\mathbf{x}_j)\} - \tau^\ell h_2^\ell(\mathbf{x}_j), \quad \forall j \in \mathcal{R}^\ell. \tag{M.7}$$

Notice that the constants (M.4) – (M.7) form the linear constraints given in Section 4. Finally, the $\mathcal{S}^\ell$ is defined explicitly by the region satisfying all the constraints. Notice that all the constraints are linear and hence the set $\mathcal{S}^\ell$ turns out to be a connected, convex, and closed polygon.

In order to keep continuing the algorithm, we set the middle points of each adjacent vertices of $\mathcal{S}^\ell$ as the next points to be updated. The solutions on the middle points can be readily updated by Theorem 1. It is essential to update the set for the middle points. Remark that each side of the polygon $\mathcal{S}^\ell$ represents a different event and a set at a middle point should be updated according to the event represented by the side. Notice also that $\mathcal{S}^\ell$ has multiple sides to be updated. The algorithm searches all the solutions by keeping continuing to obtain $\mathcal{S}^\ell$ and solutions at its vertices. The algorithm is terminated after the entire domain of $\mathcal{Q}$ is searched.

S3.3 Empty Elbow

We note that there is a possibility that $\mathcal{E}$ can be empty due to Event 1 and Event 2 and we call it *empty elbow*. If the *empty elbow* occurs then we can not apply Theorem 1. Suppose the *empty elbow* occurs at (λ^e, τ^e) then use a superscript 'e' to denote quantities obtained at (λ^e, τ^e).

It is not difficult to verify that two conditions are satisfied under the *empty elbow*: i) $\tau^e = \sum_{i\mathcal{L}^e} \omega_i / \sum_{i=1}^n \omega_i$ and ii) $\hat{\theta}_i^e$ are unique while $\hat{\theta}_0^e$ is not. In fact, $\hat{\theta}_0^e$ can be any value in the following interval,

$$[a_L, a_U] := \left[\max_{i \in \mathcal{L}^e} m_i^e, \quad \min_{i \in \mathcal{R}^e} m_i^e \right], \tag{M.8}$$

where $m_i^e = y_i \lambda^e - \sum_{j=1}^n \hat{\theta}_j^e K(\mathbf{x}_i, \mathbf{x}_j)$. Since solution path of θ_0 is continuous, the *empty elbow* can be resolved only by $\hat{\theta}_0$ touching one of the two boundaries, a_L or a_U. We regard the $\hat{\theta}_0^e$ obtained from the algorithm at (λ^e, τ^e) as an entrance to the *empty elbow* and it must be one of the a_L or a_U. Without loss of generality, we suppose $\hat{\theta}_0^e = a_L$ then the *empty elbow* is resolved by reaching another boundary a_U, which can be regarded as an exit. Therefore under the *empty elbow*, the sets can be updated as follows: Let $i_L^e = \text{argmax}_{i \in \mathcal{L}^e} m_i^e$ and $i_U^e = \text{argmin}_{i \in \mathcal{R}^e} m_i^e$, then the next $\mathcal{E}$, denoted by $\mathcal{E}^{e+1}$ is updated from the empty set to $\{i_U^e\}$. The updated $\mathcal{L}$ and $\mathcal{R}$ under *empty elbow*, denoted by $\mathcal{L}^{e+1}$ and $\mathcal{R}^{e+1}$, respectively can be accordingly obtained from the fact that the i_U^e is one from either of the two sets. In case of $\alpha_0^e = a_U$, we will have $\mathcal{E} = \{i_L^e\}$ and the other two sets updated accordingly.

Bibliography

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Zhang, T. (2002). Covering number bounds of certain regularized linear function classes, *Journal of Machine Learning Research* **2**: 527–550.