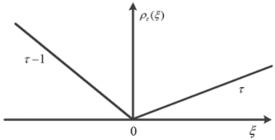


Yikun Zhang
Department of Statistics,
University of Washington
$\ast$ Part of the slides were made when I was an
Advanced Algorithmic Engineer at Trip.com

Conditional Quantile Regression

With Applications to User-Preferred Price Prediction

December 23, 2021

Table of Contents

Introduction to Our Hotel Ranking Task

A group of candidate hotels (in a searched city).

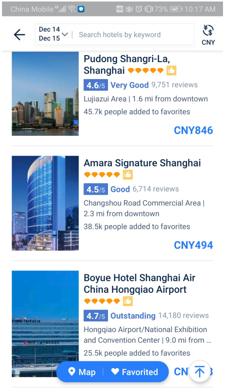
*Logistic Regression, XGBoost, Deep Neural Networks,...

Introduction to Our Hotel Ranking Task

A group of candidate hotels (in a searched city).

Ranking Algorithms* $\Longrightarrow$

A well-sorted list of hotels.



*Logistic Regression, XGBoost, Deep Neural Networks,...

Objective of the Hotel Ranking Task

Return a list of hotels with user-preferred ones placed on the top.
$\Rightarrow$ Optimizing the *conversion rate* (on hotels with high commissions).

Objective of the Hotel Ranking Task

Return a list of hotels with user-preferred ones placed on the top.
$\Rightarrow$ Optimizing the *conversion rate* (on hotels with high commissions).

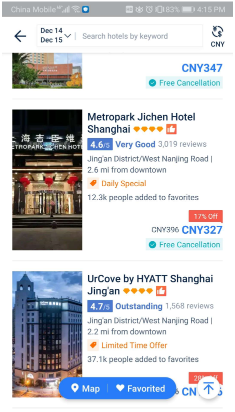
Features/Predictors: $\boldsymbol{X}_i = \left[\underbrace{V_1^{(i)}, ..., V_q^{(i)}}_{\text{Hotel Features}}, \underbrace{U_1^{(i)}, ..., U_p^{(i)}}_{\text{User Features}}\right]$ for $i = 1, ..., n$.

Responses: $Y_i \in \{0 : \text{Not Booked}, 1 : \text{Booked}\}$ for $i = 1, ..., n$.

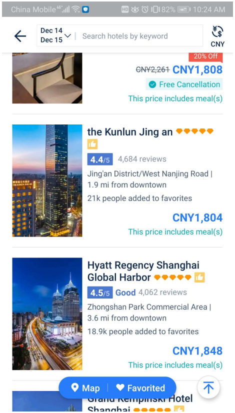
How to Identify User-Preferred Hotels?

- The prices of hotels clicked/booked by a user quantify his/her affordability.
- The price preferences of users on our platform are diverse.

Variety of User Price Preferences



(a) Users that prefer low-priced hotels



(b) Users that prefer high-priced hotels

Multimodal Nature of User Price Preferences

The price preferences varies between different groups of users on our platform.

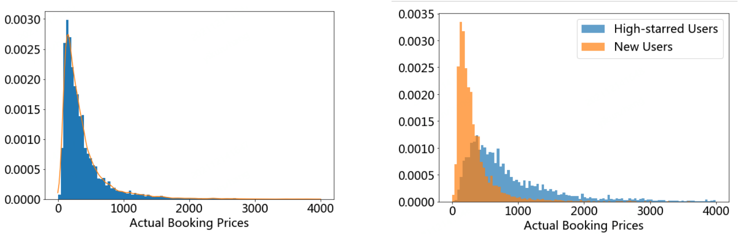


Figure 2: Overall and group-specific distributions of actual booking prices on December 6, 2021.

Main Objective: User-Preferred Hotel Price Prediction

Correctly predicting the preferred hotel prices or *price intervals* is of great significance to our hotel ranking task!

Main Objective: User-Preferred Hotel Price Prediction

Correctly predicting the preferred hotel prices or *price intervals* is of great significance to our hotel ranking task!

- Our current re-ranking mechanism relies on the predicted user-preferred prices.

Main Objective: User-Preferred Hotel Price Prediction

Correctly predicting the preferred hotel prices or *price intervals* is of great significance to our hotel ranking task!

- Our current re-ranking mechanism relies on the predicted user-preferred prices.

Mathematically, given a user $\boldsymbol{X}_i = \boldsymbol{x}_i = \left[u_1^{(i)}, ..., u_p^{(i)}\right]$, we intend to predict his/her preferred price interval

$$\left[\widehat{Q}_l(\boldsymbol{x}_i), \widehat{Q}_u(\boldsymbol{x}_i)\right].$$

Main Objective: User-Preferred Hotel Price Prediction

Correctly predicting the preferred hotel prices or *price intervals* is of great significance to our hotel ranking task!

- Our current re-ranking mechanism relies on the predicted user-preferred prices.

Mathematically, given a user $\boldsymbol{X}_i = \boldsymbol{x}_i = \left[u_1^{(i)}, ..., u_p^{(i)}\right]$, we intend to predict his/her preferred price interval

$$\left[\widehat{Q}_l(\boldsymbol{x}_i), \widehat{Q}_u(\boldsymbol{x}_i)\right].$$

Here, the features $u_j^{(i)}, j = 1, ..., p$ range from

- user behaviors (such as historical clicked/booked hotels, user IDs, etc.)
- location information (such as city IDs, average GMV in that city, etc.)

Drawback of the Current Online Model (Baseline)

Current online model: It is a weighted sum of historical booked prices, real-time clicked prices, and the specific quantile price in the searched city.

$$\text{Predicted Price} = \frac{\sum_i \omega_{\text{time}} \cdot \omega_{\text{type}} \cdot \omega_{\text{abnormal}} \cdot \omega_{\text{city}} \cdot \text{Price}_i}{\sum_i \omega_{\text{time}} \cdot \omega_{\text{type}} \cdot \omega_{\text{abnormal}} \cdot \omega_{\text{city}}}.$$

Drawback of the Current Online Model (Baseline)

Current online model: It is a weighted sum of historical booked prices, real-time clicked prices, and the specific quantile price in the searched city.

$$\text{Predicted Price} = \frac{\sum_i \omega_{\text{time}} \cdot \omega_{\text{type}} \cdot \omega_{\text{abnormal}} \cdot \omega_{\text{city}} \cdot \text{Price}_i}{\sum_i \omega_{\text{time}} \cdot \omega_{\text{type}} \cdot \omega_{\text{abnormal}} \cdot \omega_{\text{city}}}.$$

- The choices weights $\omega_{\text{time}}, \omega_{\text{type}}, \omega_{\text{abnormal}}, \omega_{\text{city}}$ are heuristic and outdated.
- The preferred price interval is symmetrically extended from the above point estimate.
- The accuracy of the current predicted prices (or price intervals) is also limited.
- ...

Our Proposed Method: Conditional Quantile Regression

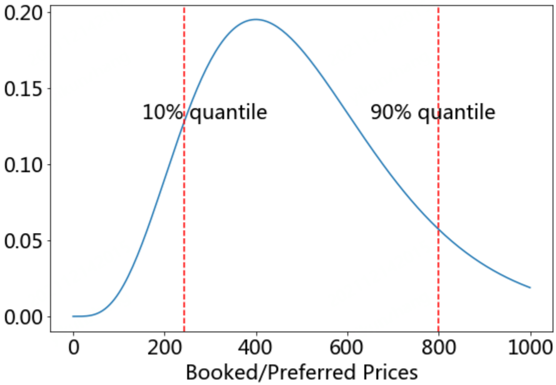


Figure 3: (Smoothed) conditional distribution of historical booked/preferred prices for a user with feature $\boldsymbol{X}_i = \boldsymbol{x}_i$. The synthetic density function (blue curve) is given by $f(u|\boldsymbol{x}_i) = \frac{1}{\Gamma(5)\cdot 100^5} \cdot u^4 \exp\left(-\frac{u}{100}\right)$.

Our Proposed Method: Conditional Quantile Regression

Given the conditional cumulative distribution function $F(y|\boldsymbol{X} = \boldsymbol{x})$ of booked prices, we pursue an interval

$$\Big[Q_{\tau}(\boldsymbol{x}),\, Q_{1-\tau}(\boldsymbol{x})\Big],$$

where $Q_{\tau}(\boldsymbol{x}) = \inf\left\{y : F(y|\boldsymbol{X} = \boldsymbol{x}) \geq \tau\right\}$ and $\tau \in (0, 1/2]$.[†]

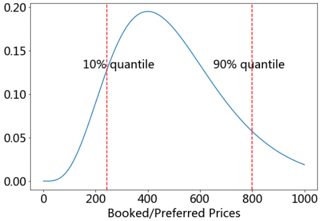


Figure 4: τ and $(1-\tau)$ quantile of $F(y|\boldsymbol{X} = \boldsymbol{x})$ with $\tau = 0.1$.

[†]Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33-50.

How to Fit the Conditional Quantile?

How to Fit the Conditional Quantile?

The conditional quantile $Q_\tau(\boldsymbol{x})$ is the solution to the following optimization problem:

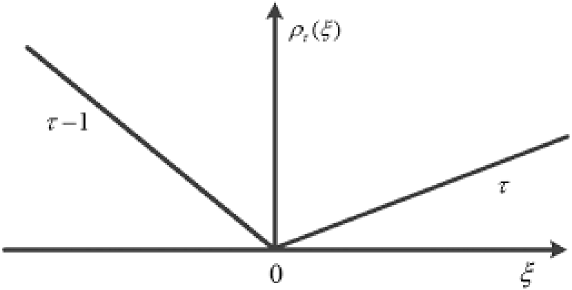
$$Q_\tau(\boldsymbol{x}) = \arg\min_q \mathbb{E}\left[\rho_\tau(Y - q) | \boldsymbol{X} = \boldsymbol{x}\right], \tag{1}$$

where

$$\rho_\tau(\xi) = \xi\left[\tau - \mathbb{1}_{\{\xi<0\}}\right] = \begin{cases} \tau\xi, & \xi \geq 0, \\ -(1-\tau)\xi, & \xi < 0 \end{cases} \tag{2}$$

is the so-called “pinball” loss (Koenker and Bassett, 1978; Firpo et al., 2009; Steinwart and Christmann, 2011).

"Pinball Loss"



Remark:

- When $\tau = 0.5$, the aforementioned optimization problem (1) recovers the absolute deviation problem.
- The loss is robust to outliers (Hampel, 1971; John, 2015).

Correctness of the (Conditional) Quantile Regression

Proposition

Given the conditional distribution function $F(y|\boldsymbol{X} = \boldsymbol{x})$,

$$Q_\tau(\boldsymbol{x}) = \inf\left\{y : F(y|\boldsymbol{X} = \boldsymbol{x}) \geq \tau\right\}$$

is the solution to (1).
More generally, given any càdlàg function $F(y)$,

$$q_\tau = \inf\left\{y : F(y) \geq \tau\right\}$$

is the solution to the unconditional quantile regression problem $\arg\min_q \mathbb{E}\left[\rho_\tau(Y - q)\right]$.

Quantile Regression in Practice

Theoretically, $Q_\tau(\boldsymbol{x}) = \arg\min_q \mathbb{E}\left[\rho_\tau(Y-q)|\boldsymbol{X}=\boldsymbol{x}\right]$.

Quantile Regression in Practice

Theoretically, $Q_\tau(\boldsymbol{x}) = \arg\min_q \mathbb{E}\left[\rho_\tau(Y-q)|\boldsymbol{X}=\boldsymbol{x}\right]$.

Practically, given the training set with clicked/booked hotel entries

$$\{(\boldsymbol{X}_i, Y_i)\} = \left\{\left(\left[U_1^{(i)}, ..., U_p^{(i)}\right], Y_i\right)\right\},$$

we solve the following empirical risk minimization (ERM) problem:

$$\widehat{Q}_\tau = \arg\min_{Q\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(Y_i - Q(\boldsymbol{X}_i)\right),$$

where $\mathcal{F}$ is the function class spanned by our (neural network) models.

Fitting the Empirical Quantiles $\widehat{Q}_\tau$ and $\widehat{Q}_{1-\tau}$

Input: $\{(\boldsymbol{X}_i, Y_i)\} = \left\{\left(\left[U_1^{(i)}, ..., U_p^{(i)}\right], Y_i\right)\right\}$, where the continuous features are standardized and discrete ones are converted to embedding vectors.

Fitting the Empirical Quantiles $\widehat{Q}_\tau$ and $\widehat{Q}_{1-\tau}$

Input: $\{(\boldsymbol{X}_i, Y_i)\} = \left\{\left(\left[U_1^{(i)}, ..., U_p^{(i)}\right], Y_i\right)\right\}$, where the continuous features are standardized and discrete ones are converted to embedding vectors.
Architecture: One shared hidden layer 512×200 with additional separate $200 \times 100 \times 1$ full-connected Relu layers.

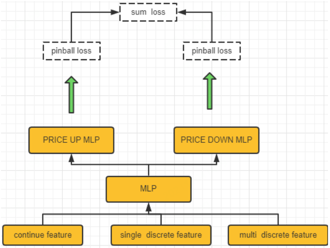


Figure 5: Double-tower architecture (image credit: Xianzhang Xiang)

Fitting the Empirical Quantiles $\widehat{Q}_\tau$ and $\widehat{Q}_{1-\tau}$

Input: $\{(\boldsymbol{X}_i, Y_i)\} = \left\{\left(\left[U_1^{(i)}, ..., U_p^{(i)}\right], Y_i\right)\right\}$, where the continuous features are standardized and discrete ones are converted to embedding vectors.
Architecture: One shared hidden layer 512×200 with additional separate $200 \times 100 \times 1$ full-connected Relu layers.

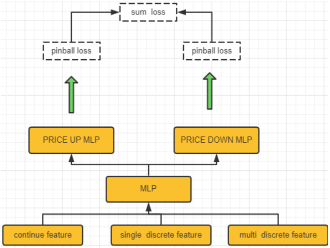


Figure 5: Double-tower architecture (image credit: Xianzhang Xiang)

Objective: $\left\{\widehat{Q}_\tau, \widehat{Q}_{1-\tau}\right\} = \underset{\{f,g\}\subset\mathcal{F}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\left[\rho_\tau\left(Y_i - f(\boldsymbol{X}_i)\right) + \rho_{1-\tau}\left(Y_i - g(\boldsymbol{X}_i)\right)\right]$

with $\tau = 0.1$.

Why do we use Relu Neural Network? (Minimax Theory)

Assume that

- the true quantile function Q_τ belongs to the Hölder class $\mathcal{H}$ or Besov space $\mathcal{B}$.
- the number of layers L satisfies $\log_2(n) \lesssim L \lesssim n^{\frac{p}{2s+p}}$.
- the maximum norm of network coefficients $\|\boldsymbol{\beta}\|_{\max} \lesssim n^{\frac{p}{2s+p}} \log n$.

Why do we use Relu Neural Network? (Minimax Theory)

Assume that

- the true quantile function Q_τ belongs to the Hölder class $\mathcal{H}$ or Besov space $\mathcal{B}$.
- the number of layers L satisfies $\log_2(n) \lesssim L \lesssim n^{\frac{p}{2s+p}}$.
- the maximum norm of network coefficients $\|\boldsymbol{\beta}\|_{\max} \lesssim n^{\frac{p}{2s+p}} \log n$.

Then,

$$\|\widehat{Q}_\tau - Q_\tau\|_{\ell_2}^2 \leq C \cdot (\log n)^2 n^{-\frac{2s}{2s+p}},$$

where s is the smoothness parameter, p is the dimension of the feature space, and n is the sample size (Schmidt-Hieber, 2020; Padilla et al., 2020).

Why do we use Relu Neural Network? (Minimax Theory)

Assume that

- the true quantile function Q_τ belongs to the Hölder class $\mathcal{H}$ or Besov space $\mathcal{B}$.
- the number of layers L satisfies $\log_2(n) \lesssim L \lesssim n^{\frac{p}{2s+p}}$.
- the maximum norm of network coefficients $\|\boldsymbol{\beta}\|_{\max} \lesssim n^{\frac{p}{2s+p}} \log n$.

Then,

$$\|\widehat{Q}_\tau - Q_\tau\|_{\ell_2}^2 \leq C \cdot (\log n)^2 n^{-\frac{2s}{2s+p}},$$

where s is the smoothness parameter, p is the dimension of the feature space, and n is the sample size (Schmidt-Hieber, 2020; Padilla et al., 2020).

Based on the nonparametric theory (Wasserman, 2006; Tsybakov, 2008), this rate of convergence is indeed *minimax* up to a log factor!

$\widehat{f}^*$ is minimax

$$\Longleftrightarrow \sup_{f\in\mathcal{H}} \mathbb{E}\left[\left(\widehat{f}^*(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\right)^2\right] = \inf_{\widehat{f}_n} \sup_{f\in\mathcal{H}} \mathbb{E}\left[\left(\widehat{f}_n(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\right)^2\right],$$

where the infimum is taken among all the estimators.

Summary of Our Proposed Model

- **Goal:** Preferred Price Interval $\left[Q_\tau(\boldsymbol{x}),\, Q_{1-\tau}(\boldsymbol{x})\right]$ with $Q_\tau(\boldsymbol{x}) = \inf\left\{y : F(y|\boldsymbol{X} = \boldsymbol{x}) \geq \tau\right\}$ and $\tau \in (0, 1/2]$.
- **Theoretical Solution:** Conditional Quantile Regression,

$$Q_\tau(\boldsymbol{x}) = \arg\min_q \mathbb{E}\left[\rho_\tau(Y - q)|\boldsymbol{X} = \boldsymbol{x}\right].$$

- **Practical Model:** Empirical Risk Minimization with Relu Networks,

$$\widehat{Q}_\tau = \arg\min_{Q\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(Y_i - Q(\boldsymbol{X}_i)\right).$$

- **Minimax Guarantee:** $\|\widehat{Q}_\tau - Q_\tau\|_{\ell_2}^2 \to 0$ as $n \to \infty$.

Other Potential Choices of Quantile Regression Models

- **Quantile Regression Forests** (Meinshausen, 2006): The random forests method has the uniform consistency in estimating the cumulative distribution function (CDF) of $Y|\boldsymbol{X} = x$.

Other Potential Choices of Quantile Regression Models

- **Quantile Regression Forests** (Meinshausen, 2006): The random forests method has the uniform consistency in estimating the cumulative distribution function (CDF) of $Y|\boldsymbol{X} = \boldsymbol{x}$.

- k**-Nearest-Neighbors (kNN) Fused Lasso** (Madrid Padilla et al., 2020; Ye and Padilla, 2021): similar to our ERM problem but with a fused lasso penalty term based on kNNs.

Other Potential Choices of Quantile Regression Models

- **Quantile Regression Forests** (Meinshausen, 2006): The random forests method has the uniform consistency in estimating the cumulative distribution function (CDF) of $Y|\boldsymbol{X} = \boldsymbol{x}$.

- k**-Nearest-Neighbors (kNN) Fused Lasso** (Madrid Padilla et al., 2020; Ye and Padilla, 2021): similar to our ERM problem but with a fused lasso penalty term based on kNNs.

- **Quadratic Programming and Reproducing Kernel Hilbert Space (RKHS) Methods** (Takeuchi et al., 2006), **Nadaraya-Watson Nonparametric Regression Estimator** (Huang and Nguyen, 2018), etc.

Evaluation Metrics

- **Coverage Accuracy**:

$$ACC(\mathcal{Y}, \widehat{\mathcal{I}}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left\{Y_i \in \left[\widehat{I}(\boldsymbol{x}_i)\right]\right\}},$$

where $\mathcal{Y} = \{Y_i\}_{i=1}^n$ is a collection of booked hotel prices and $\widehat{I}(\boldsymbol{x}_i) = \left[\widehat{Q}_\tau(\boldsymbol{x}_i), \widehat{Q}_{1-\tau}(\boldsymbol{x}_i)\right]$ is the predicted preferred price interval for the user with feature $\boldsymbol{x}_i$.

Evaluation Metrics

- **Coverage Accuracy**:

$$ACC(\mathcal{Y}, \widehat{\mathcal{I}}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left\{Y_i \in \left[\widehat{I}(\boldsymbol{x}_i)\right]\right\}},$$

where $\mathcal{Y} = \{Y_i\}_{i=1}^{n}$ is a collection of booked hotel prices and $\widehat{I}(\boldsymbol{x}_i) = \left[\widehat{Q}_\tau(\boldsymbol{x}_i), \widehat{Q}_{1-\tau}(\boldsymbol{x}_i)\right]$ is the predicted preferred price interval for the user with feature $\boldsymbol{x}_i$.

- **Average Interval Length**:

$$\text{Average Length}(\widehat{\mathcal{I}}) = \frac{1}{n}\sum_{i=1}^{n} \left|\widehat{Q}_\tau(\boldsymbol{x}_i) - \widehat{Q}_{1-\tau}(\boldsymbol{x}_i)\right|.$$

Neural Network Quantile Regression on the "My Location" Scenario

	Cov. Acc. (fh_prices)	*Cov. Acc. (order prices)*	*Average Interval Length*
Baseline Model	0.7521	0.8019	283.8624
Our NN QR	**0.8718**	**0.8593**	**233.5811**

Table 1: Comparison between our neural network quantile regression model and the current online model (baseline) on the "My Location" scenario.

Neural Network Quantile Regression on the “Main Ranking” Scenario

	Cov. Acc. (fh_prices)	*Cov. Acc. (order prices)*	*Average Interval Length*
Baseline Interval I	0.5590	0.5804	213.1197
Baseline Interval II	0.8593	0.8534	597.1224
Our NN QR (Before calibration)	0.7883	0.7351	317.5999
Our NN QR (After calibration)	**0.9268**	**0.8954**	482.3805

Table 2: Comparison between our neural network quantile regression model and the current online model (baseline) on the “Main Ranking” scenario.

- Notes: The calibration means that we extend our predicted interval as:

$$\left[\widehat{Q}_{\tau}(\boldsymbol{x}_i) - \alpha \cdot \left|\widehat{Q}_{\tau}(\boldsymbol{x}_i) - \widehat{Q}_{1-\tau}(\boldsymbol{x}_i)\right|, \widehat{Q}_{1-\tau}(\boldsymbol{x}_i) + \alpha \cdot \left|\widehat{Q}_{\tau}(\boldsymbol{x}_i) - \widehat{Q}_{1-\tau}(\boldsymbol{x}_i)\right|\right],$$

where $\alpha = 0.3 \sim 0.5$.

Discussion: Non-Crossing Property of Quantile Regression

Recall that our current optimization framework is

$$\left\{\widehat{Q}_\tau, \widehat{Q}_{1-\tau}\right\} = \underset{\{Q_\tau, Q_{1-\tau}\} \subset \mathcal{F}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\left[\rho_\tau\left(Y_i - Q_\tau(\boldsymbol{X}_i)\right) + \rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\right].$$

Discussion: Non-Crossing Property of Quantile Regression

Recall that our current optimization framework is

$$\left\{\widehat{Q}_\tau, \widehat{Q}_{1-\tau}\right\} = \mathop{\arg\min}_{\{Q_\tau, Q_{1-\tau}\}\subset\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\Big[\rho_\tau\left(Y_i - Q_\tau(\boldsymbol{X}_i)\right) + \rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\Big].$$

However, a constraint is required for the monotonicity of quantiles, *i.e.*, for any $\tau \in (0, 1/2]$, we should solve the constrained optimization problem:

$$\left\{\widehat{Q}_\tau, \widehat{Q}_{1-\tau}\right\} = \mathop{\arg\min}_{\{Q_\tau, Q_{1-\tau}\}\subset\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\Big[\rho_\tau\left(Y_i - Q_\tau(\boldsymbol{X}_i)\right) + \rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\Big]$$

subject to $Q_\tau(\boldsymbol{X}_i) \leq Q_{1-\tau}(\boldsymbol{X}_i)$ for all $i = 1, ..., n$.

Discussion: Non-Crossing Property of Quantile Regression

Recall that our current optimization framework is

$$\left\{\widehat{Q}_\tau, \widehat{Q}_{1-\tau}\right\} = \underset{\{Q_\tau, Q_{1-\tau}\}\subset\mathcal{F}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\Big[\rho_\tau\left(Y_i - Q_\tau(\boldsymbol{X}_i)\right)+\rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\Big].$$

However, a constraint is required for the monotonicity of quantiles, *i.e.*, for any $\tau \in (0, 1/2]$, we should solve the constrained optimization problem:

$$\left\{\widehat{Q}_\tau, \widehat{Q}_{1-\tau}\right\} = \underset{\{Q_\tau, Q_{1-\tau}\}\subset\mathcal{F}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\Big[\rho_\tau\left(Y_i - Q_\tau(\boldsymbol{X}_i)\right)+\rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\Big]$$

subject to $Q_\tau(\boldsymbol{X}_i) \leq Q_{1-\tau}(\boldsymbol{X}_i)$ for all $i = 1, ..., n$.

Challenges: Solving the constrained optimization problem is difficult due to the nature of stochastic gradient descent (Padilla et al., 2020).

Discussion: Solution to the Non-Crossing Constrained Quantile Regression

Feasible Approaches:

- Penalized Method: With a large $\lambda > 0$, we optimize the following problem:

$$\left\{\widehat{Q}_\tau, \widehat{Q}_{1-\tau}\right\} = \operatorname*{arg\,min}_{\{Q_\tau, Q_{1-\tau}\}\subset \mathcal{F}} \sum_{i=1}^n \Big[\rho_\tau\left(Y_i - Q_\tau(\boldsymbol{X}_i)\right) + \rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\Big]$$

$$+ \lambda \cdot \sum_{i=1}^n \mathbb{1}_{\left\{\rho_\tau(Y_i - Q_\tau(\boldsymbol{X}_i)) > \rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\right\}}.$$

Discussion: Solution to the Non-Crossing Constrained Quantile Regression

Feasible Approaches:

- Penalized Method: With a large $\lambda > 0$, we optimize the following problem:

$$\left\{\widehat{Q}_\tau, \widehat{Q}_{1-\tau}\right\} = \mathop{\arg\min}_{\{Q_\tau, Q_{1-\tau}\}\subset\mathcal{F}} \sum_{i=1}^n \Big[\rho_\tau\left(Y_i - Q_\tau(\boldsymbol{X}_i)\right) + \rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\Big]$$
$$+ \lambda \cdot \sum_{i=1}^n \mathbb{1}_{\left\{\rho_\tau(Y_i - Q_\tau(\boldsymbol{X}_i)) > \rho_{1-\tau}\left(Y_i - Q_{1-\tau}(\boldsymbol{X}_i)\right)\right\}}.$$

- Redefined Objective (Padilla et al., 2020):

$$\left\{\widehat{h}_1, \widehat{h}_2\right\} = \mathop{\arg\min}_{\{h_1, h_2\}\subset\mathcal{F}} \sum_{i=1}^n \rho_\tau\left(Y_i - h_1(\boldsymbol{X}_i)\right) + \sum_{i=1}^n \rho_{1-\tau}\left\{Y_i - h_1(\boldsymbol{X}_i) - \log\left[1 + e^{h_2(\boldsymbol{X}_i)}\right]\right\}$$

and set $\widehat{Q}_\tau(\boldsymbol{x}) = \widehat{h}_1(\boldsymbol{x})$ and $\widehat{Q}_{1-\tau}(\boldsymbol{x}) = \widehat{h}_1(\boldsymbol{x}) + \log\left[1 + e^{\widehat{h}_2(\boldsymbol{x})}\right]$.

Motivation of Our Proposed Method: Conformal Inference

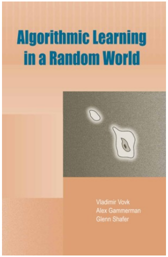
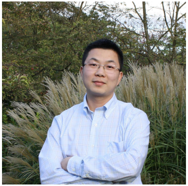
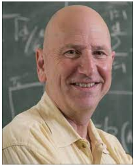


Figure 6: Algorithmic Learning in a Random World (Vovk et al., 2005).



(a) Jing Lei



(b) Larry Wasserman

(c) Emmanuel Candès

Motivation of Our Proposed Method: Conformal Inference

What is conformal prediction/inference (Vovk et al., 1999, 2005; Lei et al., 2018)?

- Given a training set $\{(\boldsymbol{X}_i, Y_i)\} \subset \mathbb{R}^p \times \mathbb{R}$ and the unknown value Y_{n+1} at a test point $\boldsymbol{X}_{n+1}$, it aims to construct a *marginal distribution-free prediction interval* $\mathcal{C}(\boldsymbol{X}_{n+1}) \subset \mathbb{R}$ such that

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1})\right) \geq 1 - \alpha$$

for some nominal miscoverage level $\alpha \in (0, 1)$.

- Notes: The $(1-\alpha)$-*confidence interval* is defined as:

$$\mathbb{P}\left(\mathbb{E}[Y|\boldsymbol{X}] \in \mathcal{C}(\boldsymbol{X})\right) \geq 1 - \alpha.$$

Classical (Split) Conformal Prediction: A Preview

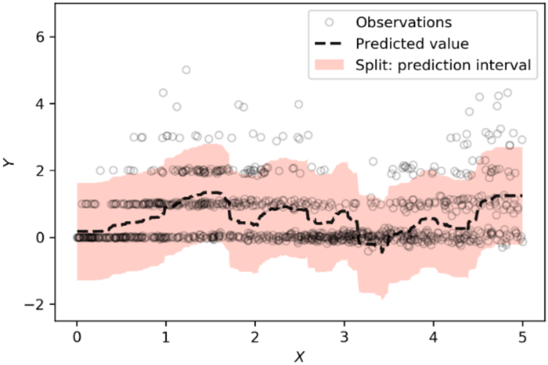


Figure 8: Classical (Split) Conformal Prediction (Average coverage: 91.4%; Average interval length: 2.91.)

Classical (Split) Conformal Prediction: Detailed Procedures

1. Split the training set $\mathcal{D} = \{(\boldsymbol{X}_i, Y_i)\} \subset \mathbb{R}^p \times \mathbb{R}$ into $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$:
   - A proper training set $\mathcal{D}_T = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}$,
   - A calibration set $\mathcal{D}_C = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_2\}$.

Classical (Split) Conformal Prediction: Detailed Procedures

1. Split the training set $\mathcal{D} = \{(\boldsymbol{X}_i, Y_i)\} \subset \mathbb{R}^p \times \mathbb{R}$ into $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$:
   - A proper training set $\mathcal{D}_T = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}$,
   - A calibration set $\mathcal{D}_C = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_2\}$.
2. Fit $\widehat{\mu}(\boldsymbol{x}) \leftarrow \mathcal{A}\left(\{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}\right)$ via any regression algorithm $\mathcal{A}$ on $\mathcal{D}_T$.

Classical (Split) Conformal Prediction: Detailed Procedures

1. Split the training set $\mathcal{D} = \{(\boldsymbol{X}_i, Y_i)\} \subset \mathbb{R}^p \times \mathbb{R}$ into $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$:
   - A proper training set $\mathcal{D}_T = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}$,
   - A calibration set $\mathcal{D}_C = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_2\}$.
2. Fit $\widehat{\mu}(\boldsymbol{x}) \leftarrow \mathcal{A}\left(\{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}\right)$ via any regression algorithm $\mathcal{A}$ on $\mathcal{D}_T$.
3. Compute the absolute residuals on $\mathcal{D}_C$ as:

$$R_i = |Y_i - \widehat{\mu}(\boldsymbol{X}_i)| \quad \text{with} \quad i \in \mathcal{I}_2.$$

Classical (Split) Conformal Prediction: Detailed Procedures

1. Split the training set $\mathcal{D} = \{(\boldsymbol{X}_i, Y_i)\} \subset \mathbb{R}^p \times \mathbb{R}$ into $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$:
   - A proper training set $\mathcal{D}_T = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}$,
   - A calibration set $\mathcal{D}_C = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_2\}$.
2. Fit $\widehat{\mu}(\boldsymbol{x}) \leftarrow \mathcal{A}\left(\{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}\right)$ via any regression algorithm $\mathcal{A}$ on $\mathcal{D}_T$.
3. Compute the absolute residuals on $\mathcal{D}_C$ as:
$$R_i = |Y_i - \widehat{\mu}(\boldsymbol{X}_i)| \quad \text{with} \quad i \in \mathcal{I}_2.$$
4. Compute the $(1-\alpha)$ empirical quantile of the absolute residuals,
$$Q_{1-\alpha}(R, \mathcal{I}_2) := (1-\alpha)\left(1 + \frac{1}{|\mathcal{I}_2|}\right)\text{-th empirical quantile of } \{R_i : i \in \mathcal{I}_2\}.$$

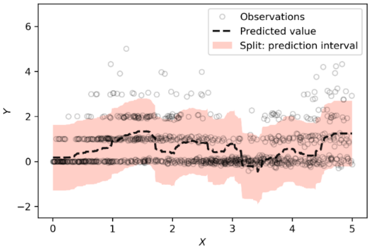
Classical (Split) Conformal Prediction: Detailed Procedures

1. Split the training set $\mathcal{D} = \{(\boldsymbol{X}_i, Y_i)\} \subset \mathbb{R}^p \times \mathbb{R}$ into $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$:
   - A proper training set $\mathcal{D}_T = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}$,
   - A calibration set $\mathcal{D}_C = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_2\}$.
2. Fit $\widehat{\mu}(\boldsymbol{x}) \leftarrow \mathcal{A}\left(\{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}\right)$ via any regression algorithm $\mathcal{A}$ on $\mathcal{D}_T$.
3. Compute the absolute residuals on $\mathcal{D}_C$ as:
$$R_i = |Y_i - \widehat{\mu}(\boldsymbol{X}_i)| \quad \text{with} \quad i \in \mathcal{I}_2.$$
4. Compute the $(1-\alpha)$ empirical quantile of the absolute residuals,
$$Q_{1-\alpha}(R, \mathcal{I}_2) := (1-\alpha)\left(1 + \frac{1}{|\mathcal{I}_2|}\right)\text{-th empirical quantile of } \{R_i : i \in \mathcal{I}_2\}.$$
5. The prediction interval at a new point $\boldsymbol{X}_{n+1}$ is given by
$$\mathcal{C}(\boldsymbol{X}_{n+1}) = \left[\widehat{\mu}(\boldsymbol{X}_{n+1}) - Q_{1-\alpha}(R, \mathcal{I}_2),\, \widehat{\mu}(\boldsymbol{X}_{n+1}) + Q_{1-\alpha}(R, \mathcal{I}_2)\right].$$

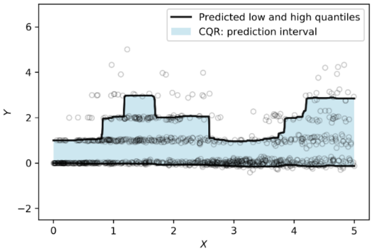
Conformalized Quantile Regression (Romano et al., 2019)

1. Split the training set $\mathcal{D} = \{(\boldsymbol{X}_i, Y_i)\} \subset \mathbb{R}^p \times \mathbb{R}$ into $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$:
   - A proper training set $\mathcal{D}_T = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}$,
   - A calibration set $\mathcal{D}_C = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_2\}$.
2. Fit $\left\{\widehat{Q}_{\alpha_{\text{low}}}, \widehat{Q}_{\alpha_{\text{high}}}\right\} \leftarrow \mathcal{A}_q\left(\{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}\right)$ via any **quantile regression** algorithm $\mathcal{A}_q$ on $\mathcal{D}_T$.
3. Compute the **conformity scores** of $\widehat{\mathcal{C}}(\boldsymbol{x}) = \left[\widehat{Q}_{\alpha_{\text{low}}}(\boldsymbol{x}), \widehat{Q}_{\alpha_{\text{high}}}(\boldsymbol{x})\right]$ on $\mathcal{D}_C$ as:
$$E_i := \max\left\{\widehat{Q}_{\alpha_{\text{low}}}(\boldsymbol{X}_i) - Y_i, Y_i - \widehat{Q}_{\alpha_{\text{high}}}(\boldsymbol{X}_i)\right\} \quad \text{with} \quad i \in \mathcal{I}_2.$$
4. Compute the $(1-\alpha)$ empirical quantile of the conformity scores,
$$Q_{1-\alpha}(E, \mathcal{I}_2) := (1-\alpha)\left(1 + \frac{1}{|\mathcal{I}_2|}\right)\text{-th empirical quantile of } \{E_i : i \in \mathcal{I}_2\}.$$
5. The prediction interval at a new point $\boldsymbol{X}_{n+1}$ is given by
$$\mathcal{C}(\boldsymbol{X}_{n+1}) = \left[\widehat{Q}_{\alpha_{\text{low}}}(\boldsymbol{X}_{n+1}) - Q_{1-\alpha}(R, \mathcal{I}_2),\ \widehat{Q}_{\alpha_{\text{high}}}(\boldsymbol{X}_{n+1}) + Q_{1-\alpha}(R, \mathcal{I}_2)\right].$$

Comparisons Between Split Conformal Prediction and Conformalized Quantile Regression



(a) Classical (Split) Conformal Prediction
(Average coverage: 91.4%; Average interval length: 2.91).



(b) Conformalized Quantile Regression
(Average coverage: 91.06%; Average interval length: 1.99).

Conclusion and Future Works

What we have done:

- We proposed a user-preferred price prediction model via (conditional) quantile regression with a Relu neural network.
- The model is well-performed based on offline evaluations.

Conclusion and Future Works

What we have done:

- We proposed a user-preferred price prediction model via (conditional) quantile regression with a Relu neural network.
- The model is well-performed based on offline evaluations.

Ongoing works:

- Handle the non-crossing properties/constraints of our model.
- Extend the user-preferred price prediction model to other scenarios and develop an unified modeling framework.

Thank You

Comments or Questions?

yikun@uw.edu

References I

S. Firpo, N. M. Fortin, and T. Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

F. R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42(6):1887–1896, 1971.

M. L. Huang and C. Nguyen. A nonparametric approach for quantile regression. *Journal of statistical distributions and applications*, 5(1):1–14, 2018.

O. O. John. Robustness of quantile regression to outliers. *American Journal of Applied Mathematics and Statistics*, 3(2):86–88, 2015.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 33–50, 1978.

J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

O. H. Madrid Padilla, J. Sharpnack, Y. Chen, and D. M. Witten. Adaptive nonparametric regression with the k-nearest neighbour fused lasso. *Biometrika*, 107(2):293–310, 2020.

References II

N. Meinshausen. Quantile regression forests. *Journal of Machine Learning Research*, 7 (35):983–999, 2006.

O. H. M. Padilla, W. Tansey, and Y. Chen. Quantile regression with deep relu networks: Estimators and minimax rates. *arXiv preprint arXiv:2010.08236*, 2020.

Y. Romano, E. Patterson, and E. Candes. Conformalized quantile regression. *Advances in Neural Information Processing Systems*, 32:3543–3553, 2019.

J. Schmidt-Hieber. Nonparametric regression using deep neural networks with relu activation function. *The Annals of Statistics*, 48(4):1875–1897, 2020.

I. Steinwart and A. Christmann. Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17(1):211–225, 2011.

I. Takeuchi, Q. V. Le, T. D. Sears, and A. J. Smola. Nonparametric quantile estimation. *Journal of Machine Learning Research*, 7(45):1231–1264, 2006.

A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 1st edition, 2008.

V. Vovk, A. Gammerman, and C. Saunders. Machine-learning applications of algorithmic randomness. In *Proceedings of International Conference on Machine Learning*, pages 444–453. Morgan Kaufmann, San Francisco, CA, 1999.

References III

V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world*. Springer Science & Business Media, 2005.

L. Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.

S. S. Ye and O. H. M. Padilla. Non-parametric quantile regression via the k-nn fused lasso. *Journal of Machine Learning Research*, 22(111):1–38, 2021.

Proof of Proposition 1.

Let $g(u) = \mathbb{E}\left[\rho_\tau(Y-u)\right]$. Some simple algebra show that

$$
\begin{aligned}
g(u) &= \int_{-\infty}^{\infty} \rho_\tau(y-u)dF(y) \\
&= \int_{u}^{\infty} \tau(y-u)dF(y) - \int_{-\infty}^{u} (1-\tau)(y-u)dF(y).
\end{aligned}
$$

Applying the Leibniz integral rule shows that

$$
g'(u) = 0 \iff -\tau \int_u^\infty dF(y) + (1-\tau)\int_{-\infty}^{u} dF(y) = F(u) - \tau = 0.
$$

Therefore, $u = q_\tau$ is the smallest point satisfying $F(u) - \tau = 0$ and will be unique when F is strictly monotonic on q_τ. $\square$