Conformalized Fairness via Quantile Regression

How to make fair prediction interval with distribution-free and non-asymptotically valid.

Chanmoo Park

January 30, 2023

Seoul National University

Problem Statement

- Fair Regression Problem

$$
\begin{aligned}
&X_i \in \mathbb{R}^p && \text{; Covariates} \\
&S_i \in [K] = \{1, \ldots, K\} && \text{; Sensitive variable (available)} \\
&Y_i \in \mathbb{R} && \text{; Target variable}
\end{aligned}
$$

- Fairness Measure

Definition: Demographic Parity (Strong DP of Wasserstein Fairness)

A prediction $g : \mathbb{R}^d \times [K] \to \mathbb{R}$ satisfies DP, if for every $s, s' \in [K]$,

$$\sup_{t \in \mathbb{R}} \left| P(g(X,S) \le t \mid S = s) - P\left(g(X,S) \le t \mid S = s'\right) \right| = 0$$

Related Works and Contributions

[1] *Fair Regression with Wasserstein Barycenter* (Evgenii et al.,2020)

- Suggested conditional mean model under DP fairness.

[2] *Conformalized Quantile Regression* (Romano et al., 2019)

- Applied ”Conformal Prediction” to the quantile predictor (Coverage Guaranteed Predcition Interval)

- This Paper : Conformalized Fairness via Quantile Regression

  - Applied **Evgenii et al.(2020)** to “Quantile Regression model”
  - Suggested “Fair CQR” : *Conformal prediction on Fairness*
  - Insight : Making ”Fair Prediction Interval” (not a point prediction) with good properties **(distribution-free, non-asymptotically valid)**

Fair Regression with Wasserstein Barycenter (Evegenii et al., 2020) I

- Consider the general regression model

$$Y = f^*(X,S) + \epsilon, \quad \text{where } f^*(X,S) = \mathbb{E}[Y|X,S]$$

Theorem: Characterization of fair optimal prediction

$$\min_{g \text{ is fair}} \mathbb{E}\left(Y - g(X,S)\right)^2 = \min_{\nu} \sum_{s \in \mathcal{S}} p_s \mathcal{W}_2^2\left(\nu_{f^*|s}, \nu\right) \tag{1}$$

where $p_s = \mathbb{P}\left(S = s\right)$ and $\mathcal{W}_2^2\left(\mu, \nu\right)$ is Wasserstein-2 distance. Moreover, minimizer of (LHS) of (1) is

$$g^*(x,s) = \left(\sum_{s' \in \mathcal{S}} p_{s'} Q_{f^*|s'}\right) \circ F_{f^*|s}\left(f^*(x,s)\right).$$

- Optimal fair predictor g^* is obtained

  by solving Wasserstein Barycenter Problem ; (RHS) of (1)

Fair Regression with Wasserstein Barycenter (Evegenii et al., 2020) II

- Optimal fair predictor g^* can be estimated by the plug-in manner

$$g^*(x,s) = \left(\sum_{s' \in \mathcal{S}} p_{s'} Q_{f^*|s'}\right) \circ F_{f^*|s}\left(f^*(x,s)\right).$$

p_s ; True Freq. of group s

$\leftarrow \hat{p_s}$; Empirical Freq. of group s

$f^*(x,s)$; Reg. func. minimizing the squared risk

$\leftarrow \hat{f}(x,s) + \epsilon$; Base estimator + jittering

$F_{f^*|s}$; CDF of $f^*(X,S)|S = s$

$\leftarrow \hat{F}_{\hat{f}|s}$; Empirical CDF of $(\hat{f}(X,S) + \epsilon)|S = s$

$Q_{f^*|s}$; Generalized inverse of F

$\leftarrow \hat{Q}_{\hat{f}|s}$; Empirical Quantile of $(\hat{f}(X,S) + \epsilon)|S = s$

$$\hat{g}(x,s) = \left(\sum_{s' \in \mathcal{S}} \hat{p}_{s'} \hat{Q}_{\hat{f}|s'}\right) \circ \hat{F}_{\hat{f}|s}(\hat{f}(x,s) + \varepsilon)$$

Conformal Prediction I

- Before dealing with
  Conformalized Quantile Regression (Romano et al., 2019)...
- **What is "Conformal" Prediction?**
  Constructing a prediction set (Not A Point Prediction)
  with **distribution-free** and **non-asymptotic** coverage guarantee.
  - *Algorithmic Learning in a Random World* (Vladimir Vovk et al., 2005).

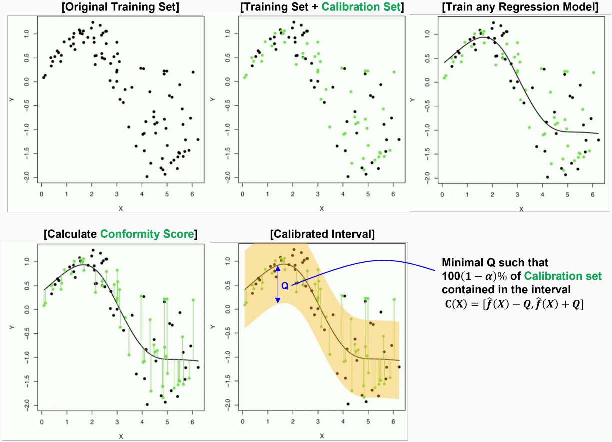
Definition: Coverage Guarantee (=Prediction Set is Valid)

For a given miscoverage rate α (for example, $\alpha = 0.1$),
a prediction set $C_n(X, S)$ has $(1-\alpha)$ Coverage Guarantee,
if for a new observation $(X_{n+1}, S_{n+1}, Y_{n+1})$,

$$P\left\{Y_{n+1} \in C_n\left(X_{n+1}, S_{n+1}\right)\right\} \geq 1-\alpha$$

- usually, C_n's are calculated
  under some distribution assumption (i.e., gaussian assumption)
  or asymptotically valid (i.e., $P\left\{Y_{test} \in C_n\left(X_{test}\right)\right\} \rightarrow 1-\alpha$).

- How to Conformalize your prediction? : simplest example

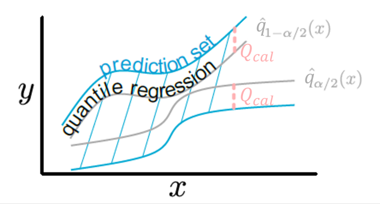


Source: *Conformal Prediction in 2022.* (Keynote speech from NeuRIPS 2022, Emmanuel, J. Candés)

- Such PI(prediction interval) $C(X) = [\hat{f}(X) - Q, \hat{f}(X) + Q]$ has distribution-free, non-asymptotic coverage property (Vovk, 2005)

Conformalized Quantile Regression(CQR) I

- Quantile Regression itself is very userful for constructing PI.
  Consider two quantile predictor $\hat{q}_{\alpha/2}(X)$, $\hat{q}_{1-\alpha/2}(X)$.



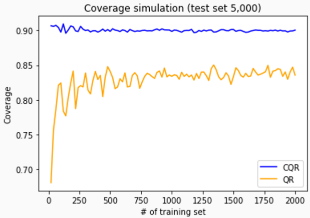
Source: A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification (Anastasios and Stephen, 2022)

  Under certain regularity conditions and for specific models, PI of $[\hat{q}_{\alpha/2}(X), \hat{q}_{1-\alpha/2}(X)]$ is asymptotically valid.

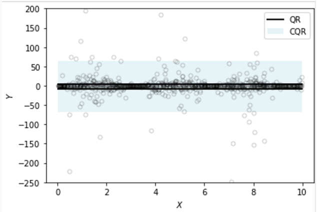
- CQR makes it non-asymptotically valid.
  $[\ \hat{q}_{train,\alpha/2}(X) - Q_{cal}, \quad \hat{q}_{train,1-\alpha/2}(X) + Q_{cal}\ ]$

- **Synthetic Data Experiments** $(Y_i \sim \mathrm{Cauchy}(0, 6\sin^2(X_i)))$

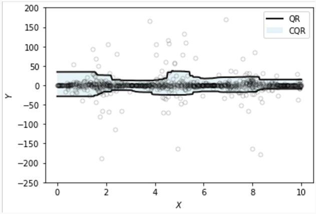
    Coverage for 5,000 test data points



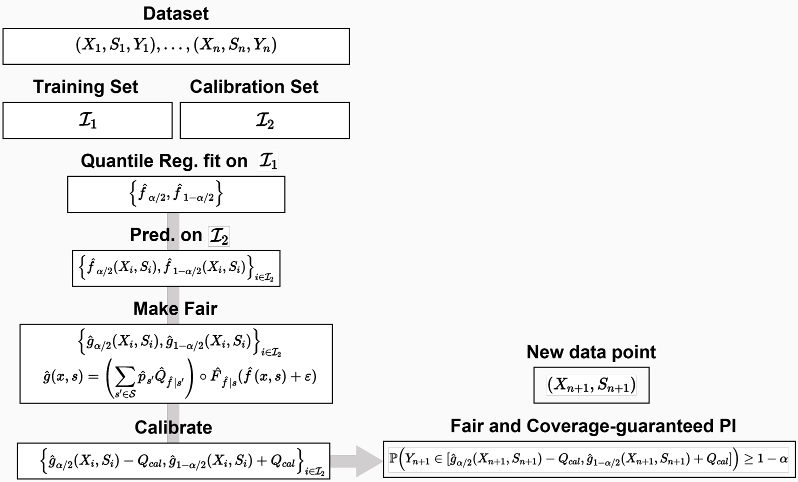
50 training set



2,000 training set



- Algorithm


Dataset
$(X_1, S_1, Y_1), \ldots, (X_n, S_n, Y_n)$
Training Set
$\mathcal{I}_1$
Calibration Set
$\mathcal{I}_2$
Quantile Reg. fit on $\mathcal{I}_1$
$\left\{\hat{f}_{\alpha/2}, \hat{f}_{1-\alpha/2}\right\}$
Pred. on $\mathcal{I}_2$
$\left\{\hat{f}_{\alpha/2}(X_i, S_i), \hat{f}_{1-\alpha/2}(X_i, S_i)\right\}_{i \in \mathcal{I}_2}$
Make Fair
$\left\{\hat{g}_{\alpha/2}(X_i, S_i), \hat{g}_{1-\alpha/2}(X_i, S_i)\right\}_{i \in \mathcal{I}_2}$
$\hat{g}(x, s) = \left(\sum_{s' \in \mathcal{S}} \hat{p}_{s'} \hat{Q}_{\hat{f}|s'}\right) \circ \hat{F}_{\hat{f}|s}(\hat{f}(x, s) + \varepsilon)$
Calibrate
$\left\{\hat{g}_{\alpha/2}(X_i, S_i) - Q_{cal}, \hat{g}_{1-\alpha/2}(X_i, S_i) + Q_{cal}\right\}_{i \in \mathcal{I}_2}$
New data point
(X_{n+1}, S_{n+1})
Fair and Coverage-guaranteed PI
$\mathbb{P}\Big(Y_{n+1} \in [\hat{g}_{\alpha/2}(X_{n+1}, S_{n+1}) - Q_{cal}, \hat{g}_{1-\alpha/2}(X_{n+1}, S_{n+1}) + Q_{cal}]\Big) \geq 1 - \alpha$

Conformalized Fairness via Quantile Regression II

- Numerical Experiments

  Law School (LAW) : GPA prediction, S is gender

  Community & Crime (CRIME) : Crime prediction, S is race

  MEPS 2016 (MEPS) : Health care prediction, S is race

  Government Salary (GOV) : Salary prediction, S is race

	LAW				CRIME			
	Coverage	Length	KS(lo)	KS(hi)	Coverage	Length	KS(lo)	KS(hi)
Ln-CQR	90.16±0.47	0.46±.004	0.39±0.03	0.11±0.02	90.22±1.88	1.30±0.05	0.62±0.06	0.53±0.06
Ln-CFQP	90.02±0.51	0.46±.004	0.02±0.01	0.02±0.01	90.44±1.84	1.64±0.05	0.11±0.03	0.12±0.04
RF-CQR	90.25±0.55	0.39±.005	0.20±0.02	0.15±0.02	90.27±1.66	1.15±0.03	0.64±0.05	0.59±0.05
RF-CFQP	90.11±0.48	0.38±.004	0.02±.008	0.02±.009	90.34±1.84	1.54±0.04	0.12±0.04	0.12±0.03
NN-CQR	90.00±0.50	0.40±0.02	0.41±0.07	0.18±0.05	90.01±1.89	1.16±0.05	0.70±0.05	0.63±0.06
NN-CFQP	90.01±0.51	0.39±0.01	0.02±.009	0.03±.009	89.95±1.62	1.54±0.12	0.12±0.04	0.12±0.03
	MEPS				GOV			
	Coverage	Length	KS (lo)	KS(hi)	Coverage	Length	KS (lo)	KS(hi)
Ln-CQR	89.92±0.66	0.66±0.01	0.09±0.03	0.33±0.05	90.00±0.19	0.79±.002	0.26±.014	0.44±0.02
Ln-CFQP	89.99±0.69	0.66±0.01	0.03±0.01	0.03±0.01	90.02±0.19	0.78±.002	0.05±0.01	0.04±0.01
RF-CQR	90.07±0.65	0.38±.009	0.19±0.02	0.30±0.03	90.03±0.17	0.61±.002	0.29±0.01	0.28±0.02
RF-CFQP	90.38±0.60	0.39±0.01	0.02±0.01	0.03±0.01	90.03±0.17	0.62±.002	0.05±0.01	0.04±0.01
NN-CQR	89.95±0.68	0.37±0.04	0.24±0.09	0.37± 0.06	90.01±0.19	0.58±0.01	0.28±0.03	0.32±0.04
NN-CFQP	89.97±0.61	0.37±0.04	0.03±0.01	0.04±0.01	90.01±0.18	0.59±0.01	0.05±0.01	0.05±0.01

Conclusion

- Constructed ”Fair” Prediction Interval in Regression Problem

- Not only to achieve **Demographic Parity**
  by Evegenii et al. (2020)

- But also to achieve **Distribution-free** & **Non-Asymptotically valid**
  by Romano et al. (2019)