Notation

- $[K] := \{1, 2, \ldots, k\}$ for given $K \in \mathbb{N}$
- $|\mathcal{S}|$: the cardinality for a finite set $\mathcal{S}$
- $\mathbb{1}\{\cdot\}$: indicator function
- $(X_1, S_1, Y_1), \ldots, (X_n, S_n, Y_n)$: observed triplet where (X_i, S_i, Y_i) is denoted by a rantom variable $Z_i \in \mathbb{R}^p \times [K] \times \mathbb{R}$
- $\mathcal{I}_1, \mathcal{I}_2$: training set and calibration set respectively

Notation

- $q_\alpha : \mathbb{R}^p \times [K] \to \mathbb{R}$: quantile predictor
- $\nu_{q_\alpha|s}$: distribution of $q_\alpha(X,S)|S=s$
- $F_{q_\alpha|s} : \mathbb{R} \times [K] \to [0,1]$: CDF of $\nu_{q_\alpha|s}$

$$F_{q_\alpha|s} = P(q_\alpha(X,S) \leq t | S = s)$$

- $Q_{q_\alpha|s} = F^{-1}_{q_\alpha|s} : [0,1] \to \mathbb{R}$: quantile function

$$Q_{q_\alpha|s} = \inf\{y \in \mathbb{R} : F_{q_\alpha|s}(y) \geq t\} \text{ with } Q_{q_\alpha|s}(0) = Q_{q_\alpha|s}(0+)$$

- Goal of conformal prediction is to construct a marginal distribution-free prediction band $C(X_{n+1}) \subseteq \mathbb{R}$ which satisfies

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

  with miscoverage rate α

CQR - Conformalized Quantile Regression

- Conformalized Quantile Regression(CQR) aims to construct a prediction band which is adaptive to heteroskedascity
- Given any quantile regression algorithm $\mathcal{Q}$, fit two conditional quantile functions $\hat{q}_{\alpha_{lo}}$ and $\hat{q}_{\alpha_{hi}}$ on trainig set $\mathcal{I}_1$. (ex: for $\alpha = 0.05$, $\{\alpha_{lo}, \alpha_{hi}\} = \{\alpha_{0.025}, \alpha_{0.975}\}$)

$$\{\hat{q}_{\alpha_{lo}}, \hat{q}_{\alpha_{hi}}\} \leftarrow \mathcal{Q}\left(\left\{\left(\tilde{X}_i, Y_i\right) : i \in \mathcal{I}_1\right\}\right)$$

- The conformity scores $E = \{E_k : k \in \mathcal{I}_2\}$ are computed as

$$E_k = \max\{\hat{q}_{\alpha_{lo}}(\tilde{X}_k) - Y_k, Y_k - \hat{q}_{\alpha_{hi}}(\tilde{X}_k)\} \text{ for } k \in \mathcal{I}_2$$

- Let $Q_{1-\alpha}(E)$ be $(1-\alpha)(1+1/|\mathcal{I}_2|)$-th empirical quantile of E
- By CQR, the prediction interval for Y_{n+1} is constructed as follow;

$$C\left(\tilde{X}_{n+1}\right) = \left[\hat{q}_{\alpha_{lo}}\left(\tilde{X}_{n+1}\right) - Q_{1-\alpha}(E), \hat{q}_{\alpha_{hi}}\left(\tilde{X}_{n+1}\right) + Q_{1-\alpha}(E)\right]$$

Demographic Parity

- For an arbitrary prediction $g : \mathbb{R}^d \times [K] \to \mathbb{R}$ satisfies demographic parity under a distribution P over (X, S, Y), if $g(X, S)$ is statistically independent of the sensitive attribute S.
- That is, for every $s, s' \in [K]$,

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

CFQP - Conformal Fair Quantile Prediction

- Target is to construct DP fairness constrained prediction intervals by using fair quantile regression algorithm

Proposition 1 (Fair optimal prediction [12]). Assume, for each $s \in [K]$, that the univariate measure $\nu_{q_\alpha|s}$ has a density and let $p_s = P(S = s)$. Then,

$$\min_{g_\alpha \text{ is fair}} E\left(q_\alpha(X,S) - g_\alpha(X,S)\right)^2 = \min_{\nu} \sum_{s \in [K]} p_s \mathcal{W}_2^2\left(\nu_{q_\alpha|s}, \nu\right). \quad (5)$$

Moreover, if g_α and ν solve the l.h.s. and the r.h.s. problems respectively, then $\nu = \nu_{g_\alpha}$ and specifically,

$$g_\alpha(x,s) = \sum_{s' \in [K]} p_{s'} Q_{q_\alpha|s'} \circ F_{q_\alpha|s} \circ q_\alpha(x,s). \quad (6)$$

CFQP - Conformal Fair Quantile Prediction

- Fit an arbitrary quantile regression algorithm $\mathcal{Q}$ on $\mathcal{I}_1$

$$\{\hat{q}_{\alpha_{lo}}, \hat{q}_{\alpha_{hi}}\} \leftarrow \mathcal{Q}\left(\left\{\left(\tilde{X}_i, Y_i\right) : i \in \mathcal{I}_1\right\}\right)$$

- Transform $\hat{q}_\alpha$ into $\hat{g}_\alpha$ by using proposition 1 to compute fair conformity scores $E^f = \{E_k^f : k \in \mathcal{I}_2\}$

$$E_k^f = \max\{\hat{g}_{\alpha_{lo}}(\tilde{X}_k) - Y_k, Y_k - \hat{g}_{\alpha_{hi}}(\tilde{X}_k)\}$$

- Then the fair prediction interval for Y_{n+1} is constructed as

$$C(\tilde{X}_{n+1}) = \left[\hat{g}_{\alpha_{lo}}(\tilde{X}_{n+1}) - Q_{1-\alpha}(E^f), \hat{g}_{\alpha_{hi}}(\tilde{X}_{n+1}) + Q_{1-\alpha}(E^f)\right]$$

CFQP - Conformal Fair Quantile Prediction

- For fair conformity scores $E^f = \{E_i^f : i \in \mathcal{I}_2\}$

$$E_i^f = \max\{\hat{g}_{\alpha_{lo},i} - Y_i, Y_i - \hat{g}_{\alpha_{hi},i}\}$$

with $\hat{q}^s_{2,\alpha,i}$ is calculated from fitted quantile regression $\mathcal{Q}$

$$\hat{g}_{\alpha,i} = \left(\sum_{s' \in [k]} \hat{p}_{s'} \hat{Q}_{2,q_\alpha|s'}\right) \circ \hat{F}_{q_\alpha|s} \circ \tilde{q}^s_{2,\alpha,i}$$

$$\tilde{q}^s_{2,\alpha,i} = \hat{q}^s_{2,\alpha,i} + U_i^s([-\sigma,\sigma]) \quad \forall i \in \mathcal{I}_2^s, s \in [K]$$

$$\hat{F}_{q_\alpha|s}(t) = \frac{1}{|\mathcal{I}_2^s|} \sum_{i \in |\mathcal{I}_2^s|} \mathbb{1}\left\{\tilde{q}^s_{2,\alpha,i} \leq t\right\}$$

$$\hat{Q}_{2,q_\alpha|s}(t) = \int_0^1 \hat{F}^{-1}_{q_\alpha|s}(v) K_h(t-v)dv \quad t \in (0,1)$$

with some smoothing kernel function K_h

CFQP - Conformal Fair Quantile Prediction

- For a new data point $\tilde{X}_{n+1} = (x, s)$ and $\alpha \in \{\alpha_{lo}, \alpha_{hi}\}$, the Fair prediction interval for Y_{n+1} is constructed as

$$C(\tilde{X}_{n+1}) = \left[\hat{g}_{\alpha_{lo}}(x, s) - Q_{1-\alpha}(E^f), \hat{g}_{\alpha_{hi}}(x, s) + Q_{1-\alpha}(E^f)\right]$$

where

$$\hat{g}_\alpha(x, s) = \left(\sum_{s' \in [K]} \hat{p}_{s'} \hat{Q}_{2,q_\alpha|s'}\right) \circ \hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x, s) \ \forall \alpha \in \{\alpha_{lo}, \alpha_{hi}\}$$

$$\tilde{q}^s_{1,\alpha,i} = \hat{q}^s_{1,\alpha,i} + U^s_i([-\sigma, \sigma]) \quad \forall i \in \mathcal{I}^s_1 \text{ and } \tilde{q}_\alpha(x, s) = \hat{q}_\alpha(x, s) + U([-\sigma, \sigma])$$

$$\hat{F}_{1,q_\alpha|s}(t) = \frac{1}{|\mathcal{I}^s_1| + 1}\left(\sum_{i \in |\mathcal{I}^s_1|} \mathbb{1}(\tilde{q}^s_{1,\alpha,i} < t) + U([0,1])\left(1 + \sum_{i \in |\mathcal{I}^s_1|} \mathbb{1}(\tilde{q}^s_{1,\alpha,i} < t)\right)\right)$$

CFQP - Conformal Fair Quantile Prediction

Theorem 1 (Prediction coverage guarantee). If $(\bar{X}_i, Y_i), i = 1, \ldots, n+1$ are exchangeable, then the prediction interval $C(\bar{X}_{n+1})$ constructed by the split CFQP algorithm satisfies

$$P\{Y_{n+1} \in C(\bar{X}_{n+1})\} \geq 1 - \alpha.$$

Moreover, if the conformity scores E_i are almost surely distinct, the prediction interval is nearly exactly calibrated,

$$P\{Y_{n+1} \in C(\bar{X}_{n+1})\} \leq 1 - \alpha + 1/(|\mathcal{I}_2| + 1).$$

Theorem 2 (Demographic parity guarantee). For any joint distribution P of (X, S, Y), any $\sigma > 0$, as well as the base quantile estimator $\hat{q}_\alpha : \mathbb{R}^p \times [K] \to \mathbb{R}$ constructed on labeled data, the estimator $\hat{g}_\alpha$ defined in Eq. (12) satisfies

$$(\hat{g}_\alpha(X, S) \mid S = s) \stackrel{d}{=} (\hat{g}_\alpha(X, S) \mid S = s') \quad \forall s, s' \in [K].$$

CFQP - Conformal Fair Quantile Prediction

Algorithm 1 Split Conformal Fair Prediction (CFQP)

Input: $\mathcal{D} = \{(X_i, S_i, Y_i)\}_{i=1}^n$; miscoverage level $\alpha \in (0, 1)$; quantile regression algorithm $\mathcal{Q}$.

```
Randomly split [n] into disjoint proper training and calibration indices I_1, I_2.
Fit two conditional quantile functions on the training set {q̂_{α_lo}, q̂_{α_hi}} ← Q({(X_i, S_i, Y_i), i ∈ I_1}).
Call functional Synchronization (Algorithm 2) to calculate {ĝ_{α_lo}, ĝ_{α_hi}} for each i ∈ I_2.
Compute E_i ← max {ĝ_{α_lo}(X_i) − Y_i, Y_i − ĝ_{α_hi}(X_i)} for ∀ i ∈ I_2.
Compute Q_{1−α}(E, I_2) ← (1 − α)(1 + 1/|I_2|)-th empirical quantile of {E_i : i ∈ I_2}.
For a new test point (x, s), compute {ĝ_{α_lo}(x, s), ĝ_{α_hi}(x, s)} through Algorithm 2
```

Output: Fair prediction interval $C(x, s) = [\hat{g}_{\alpha_{lo}}(x, s) - Q_{1-\alpha}(E, \mathcal{I}_2), \hat{g}_{\alpha_{hi}}(x, s) + Q_{1-\alpha}(E, \mathcal{I}_2)]$ for $(X_{n+1}, S_{n+1}) = (x, s)$.

Algorithm 2 Functional Synchronization

Input: Calibration set $\{(X_i, S_i)\}_{i \in \mathcal{I}_2}$ or new point (x, s); base quantile estimator $\mathcal{Q}$; slack parameter σ; training set $\{(X_i, S_i)\}_{i \in \mathcal{I}_1}$;

```
if Calibration set {(X_i, S_i)}_{i∈I_2} then
    for α ∈ {α_lo, α_hi} do
        {q̃_α(X_i, S_i)} ← {q_α(X_i, S_i) + U([−σ, σ])}_{i∈I_2}        ▷ U([−σ, σ]) are used for tie-breaking
        for s' ∈ [K] do
            Compute F̂_{q_α|s'}(t), and F̂^{−1}_{2,q_α|s'}(t) by Eq. (7) and (8).
            Obtain ĝ_α(X_i, S_i) ← Σ_{s'=1}^K p̂_{s'} F̂^{−1}_{2,q_α|s'} ∘ F̂_{q_α|s'} ∘ q̃_α(X_i, S_i), ∀i ∈ I_2
        end for
    end for
else if New test point (x, s) then
    for α ∈ {α_lo, α_hi} do
        {q̃^s_{1,α}} ← {q̂^s_α + U([−σ, σ])}_{i∈I^s_1} and q̃_α(x, s) ← q_α(x, s) + U([−σ, σ])
        Compute ĝ_α(x, s) ← Σ_{s'=1}^K p̂_{s'} F̂^{−1}_{2,q_α|s'} ∘ F̂_{1,q_α|s} ∘ q̃_α(x, s) by Eq. (8) and (7)
    end for
end if
```

Output: fair quantile prediction $\hat{g}_\alpha$ for calibration set or new test point (x, s).

Experiments

	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

Table 1: Results reported on test set of 200 repeated experiments with $\alpha = 0.1$. CQR refers to the conformalized quantile regression in [34]. Ln, RF, and NN denote the linear, random forest, and neural network quantile regression models. Our methods are shown in bold.

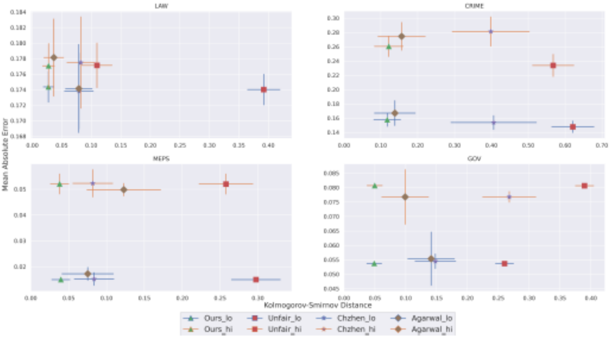

Figure 2: Results for estimating the lower (α_{lo}) and upper (α_{hi}) quantiles using some state-of-the-art DP-fairness requirement methods on all the datasets. 'Unfair', 'Chzhen', and 'Agarwal' stand for the linear quantile model without fairness adjustment, barycenter method [12] and reduction-based algorithm [1] respectively. We present the MAE and KS of lower and upper quantile estimations. A Linear quantile model is implemented in this comparison.

Algorithm 1 EM Algorithm for Fair GMM

while $Q_{fair}(\Theta; \Theta^{[t+1]}, \lambda, \omega) - Q_{fair}(\Theta; \Theta^{[t]}, \lambda, \omega) > \delta$ **do**
  E-step: Compute $Q_{fair}(\Theta; \Theta^{[t]}, \lambda, \omega)$
  M-step:
  repeat
    while Not converged and $r < R$ **do**
      $\boldsymbol{\mu}_{(r+1)} \leftarrow \boldsymbol{\mu}_{(r)} + \epsilon \frac{\partial Q_{fair}(\Theta;\Theta^{[t]},\lambda,\omega)}{\partial \boldsymbol{\mu}}$
      $\boldsymbol{\eta}_{(r+1)} \leftarrow \boldsymbol{\eta}_{(r)} + \epsilon \frac{\partial Q_{fair}(\Theta;\Theta^{[t]},\lambda,\omega)}{\partial \boldsymbol{\eta}}$
      $r \leftarrow r + 1$
    until converges
  end while
  $t \leftarrow t + 1$
end while

CFQP - Conformal Fair Quantile Prediction

- Synthetic Data
  - 400 datapoints
  - sensitive ratio: 1:1
  - sensitive centers = [(1, 1), (2.1, 1)], [(1, 7), (2.1, 7)]

Synthetic_Data.PNG

Synthetic_Assignment.PNG