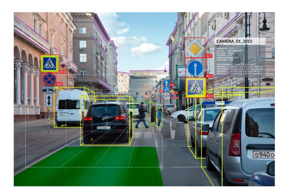
We propose a new algorithm for **conformal inference** under **covariate shift.**

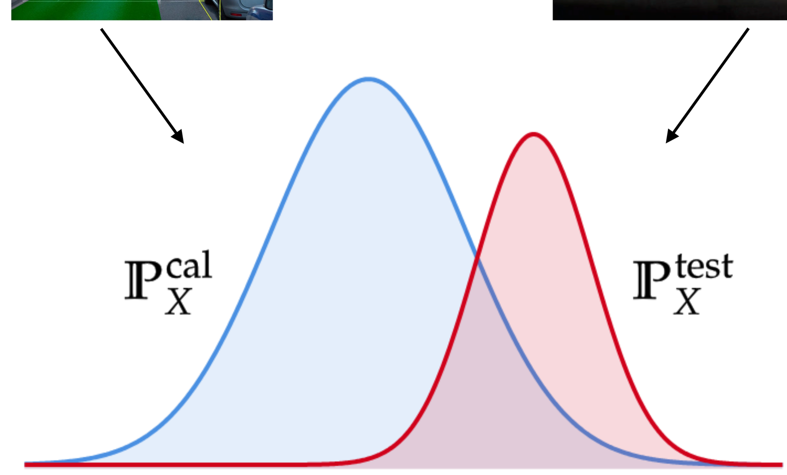
CP: Calibration data $\longrightarrow$ Threshold $h : \mathcal{X} \to \mathcal{Y}$ $\longrightarrow$ Prediction set $C(x) = \{y : S(x,y) \le h(x)\}$

Goal: $\mathbb{P}(Y_{\text{test}} \in \hat{C}(X_{\text{test}})) \ge 1 - \alpha$ (marginal test coverage)

Setting: *covariate shift* in *high (input) dim* from calibration → test

Likelihood-ratio: $r(x) = d\mathbb{P}_X^{\text{test}} / d\mathbb{P}_X^{\text{cal}}(x)$





Quantile Regression: given affine class $\mathcal{H}$,

$$h^* = \arg\min_{h \in \mathcal{H}} \mathbb{E}_{\text{cal}}[\ell_\alpha(h(X), S(X,Y))].$$

(#1) valid coverage if $r \in \mathcal{H}$

(#2) valid coverage if $h^* = q_{1-\alpha}$ (quantile fn)

Q: Interpolate between r and $q_{1-\alpha}$, without high-d estimation?

Conformal Inference under High-Dimensional Covariate Shifts via Likelihood-Ratio Regularization

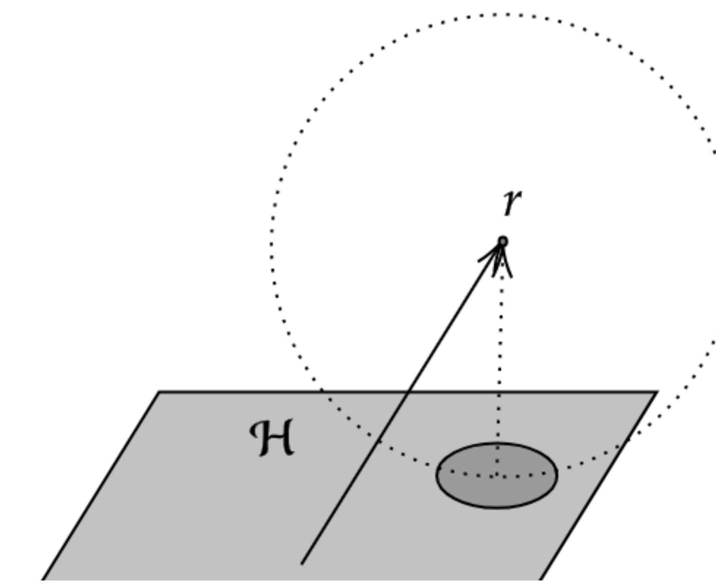
Sunay Joshi*, Shayan Kiyani*,
George J. Pappas, Edgar Dobriban, Hamed Hassani

NEURAL INFORMATION PROCESSING SYSTEMS

Penn UNIVERSITY of PENNSYLVANIA

Our Method: Likelihood-Ratio Regularization

Idea: craft novel regularization to encourage $h^* \approx r$



Step 1: Constrain $h \approx r$. $\min_{h \in \mathcal{H}, \gamma \in \mathbb{R}} \mathbb{E}_{\text{cal}}[\ell_\alpha(\gamma h, S)]$ s.t. $\mathbb{E}_{\text{cal}}[(h - r)^2)] \le \varepsilon$

Step 2: Expand. $\mathbb{E}_{\text{cal}}[(h-r)^2)] = \mathbb{E}_{\text{cal}}[h^2] + \underbrace{\mathbb{E}_{\text{test}}[-2h]}_{\text{change-of-measure}} + \underbrace{\mathbb{E}_{\text{cal}}[r^2]}_{\text{constant}}$

Step 3: Replace w/ regularizer. Drop constant: $\boxed{\mathcal{R}(h) = \mathbb{E}_{\text{cal}}[h^2] + \mathbb{E}_{\text{test}}[-2h]}$

Estimable *without plugging-in* $\hat{r}$!

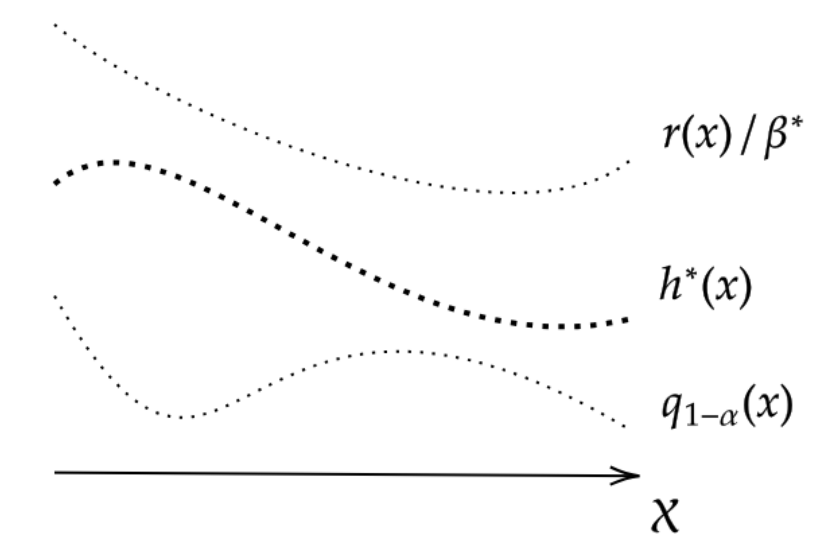
Reparametrize: $(h, \gamma) \mapsto (\beta h, 1/\beta)$:

Our Algorithm: LR-QR

1. Compute $(h^*, \beta^*) = \arg\min_{h \in \mathcal{H}, \beta \in \mathbb{R}} L_\lambda(h, \beta)$, where

$$L_\lambda(h, \beta) = \mathbb{E}_{\text{cal}}[\ell_\alpha(h, S)] + \lambda \mathcal{R}(\beta h)$$

2. Return $C(x) = \{y \in \mathcal{Y} : S(x,y) \le h^*(x)\}$



LR-QR interpolates between r and $q_{1-\alpha}$.

Learning Theory

In finite samples, PAC-style *coverage lower bound:*

Thm. With high probability, the ERM $(\hat{h}, \hat{\beta})$ with regularization strength λ obeys:

$$\mathbb{P}_{\text{test}}(Y \in \hat{C}(X)) \ge (1-\alpha) - (1-\alpha)(\text{approximation error of } \mathcal{H}) - o(1)\,.$$

Experiments

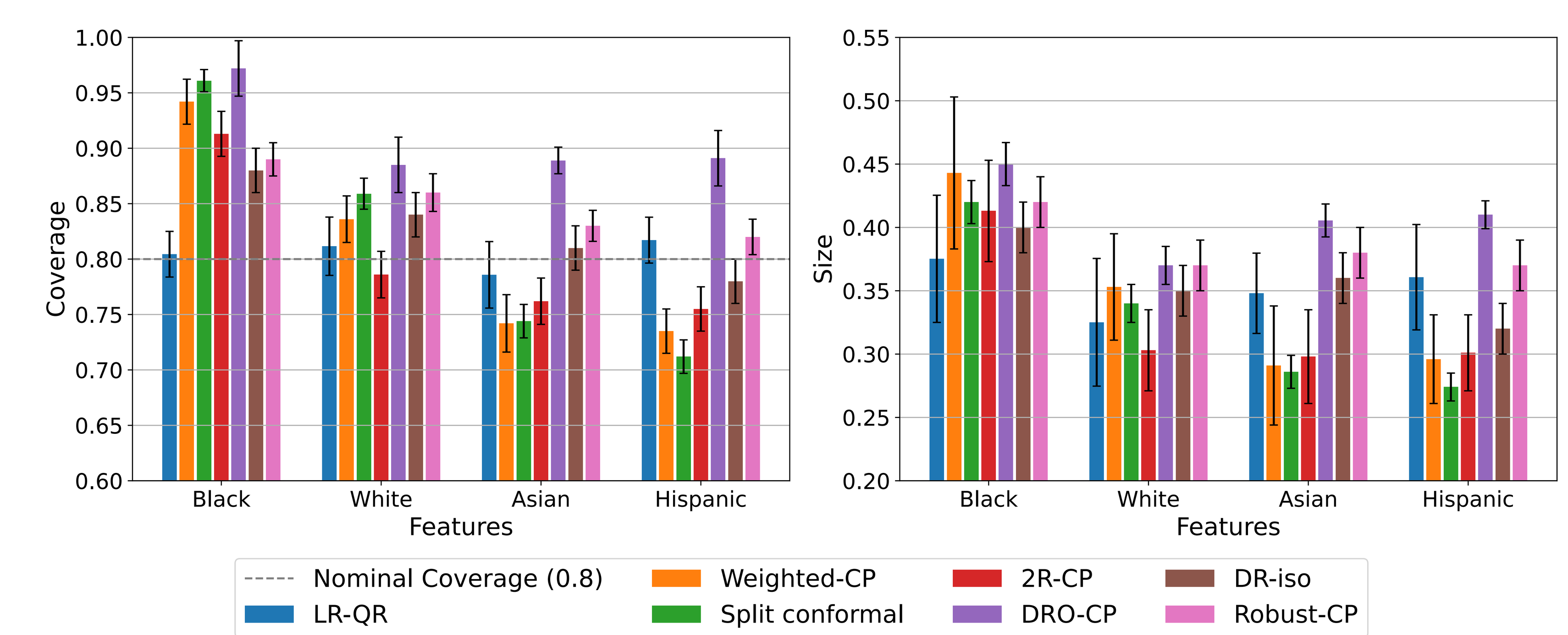


Figure 1: (Left) Coverage. (Right) Average prediction set size on the Communities and Crime dataset.

Table 1: Comparison on the MMLU benchmark (mean ± std).

Metric	Nominal	LR-QR	DRO	WCP
Coverage (%) Set Size	90.0 ± 0.0 –	89.6 ± 1.2 3.38 ± 0.15	99.7 ± 0.3 3.92 ± 0.20	86.5 ± 1.5 3.31 ± 0.12
Metric	**SCP**	**DR-iso**	**Robust-CP**	**2R-CP**
Coverage (%) Set Size	78.1 ± 2.1 2.60 ± 0.10	96.3 ± 0.6 3.64 ± 0.18	95.8 ± 0.7 3.56 ± 0.14	96.9 ± 0.5 3.80 ± 0.17