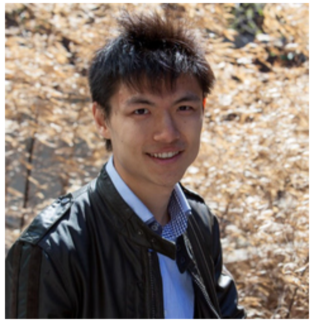
Fair Regression:
Quantitative Definitions and Reduction-Based Algorithms

Steven Wu
(University of Minnesota)
Joint work with: Alekh Agarwal and Miro Dudík
(Microsoft Research)





Problem setting

- Distribution D over examples: (X, A, Y)
  - X: feature vector
  - A: discrete protected attribute (e.g. racial groups, gender)
  - $Y \in [0, 1]$: real-valued label (e.g. risk score, recidivism rate)

- Prediction task: given loss function ℓ (e.g. square loss, logistic loss)

  find a predictor $f \in F$ to minimize $E_D[\, \ell(Y, f(X)]$

- ℓ is 1-Lipschitz:

$$|\ell(y, u) - \ell(y', u')| \leq |y - y'| + |u - u'|$$

Fairness notion: Statistical Parity

- Statistical parity (SP): $f(X)$ is independent of protected attribute A

$$P[\, f(X) \geq z \,|A = a \,] = P[\, f(X) \geq z \,]$$

for all groups a and $z \in [0, 1]$

- Implies any thresholding of $f(X)$ is fair!
- Motivated by practice of affirmative action as well as four-fifths rule

Fairness notion: Bounded Group Loss

- Bounded group loss (BGL): bounded group loss at level η

$$E_D[\ell(Y, f(X))|A = a] \leq \eta$$

for all groups a.

- Enforces minimum prediction quality for each group
- Diagnostic to detect groups requiring further data collection, better features, etc.
- Similar to minmax fairness

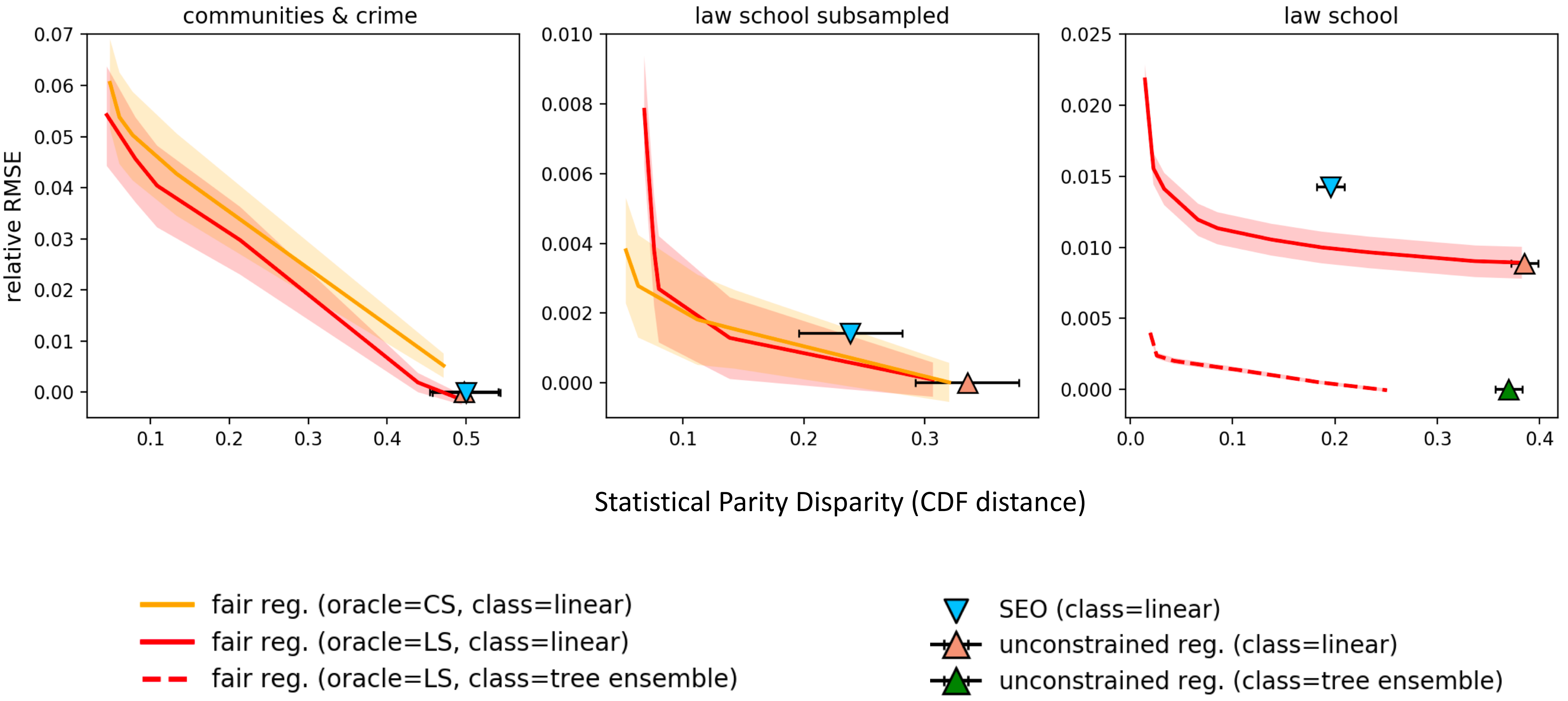
Main results

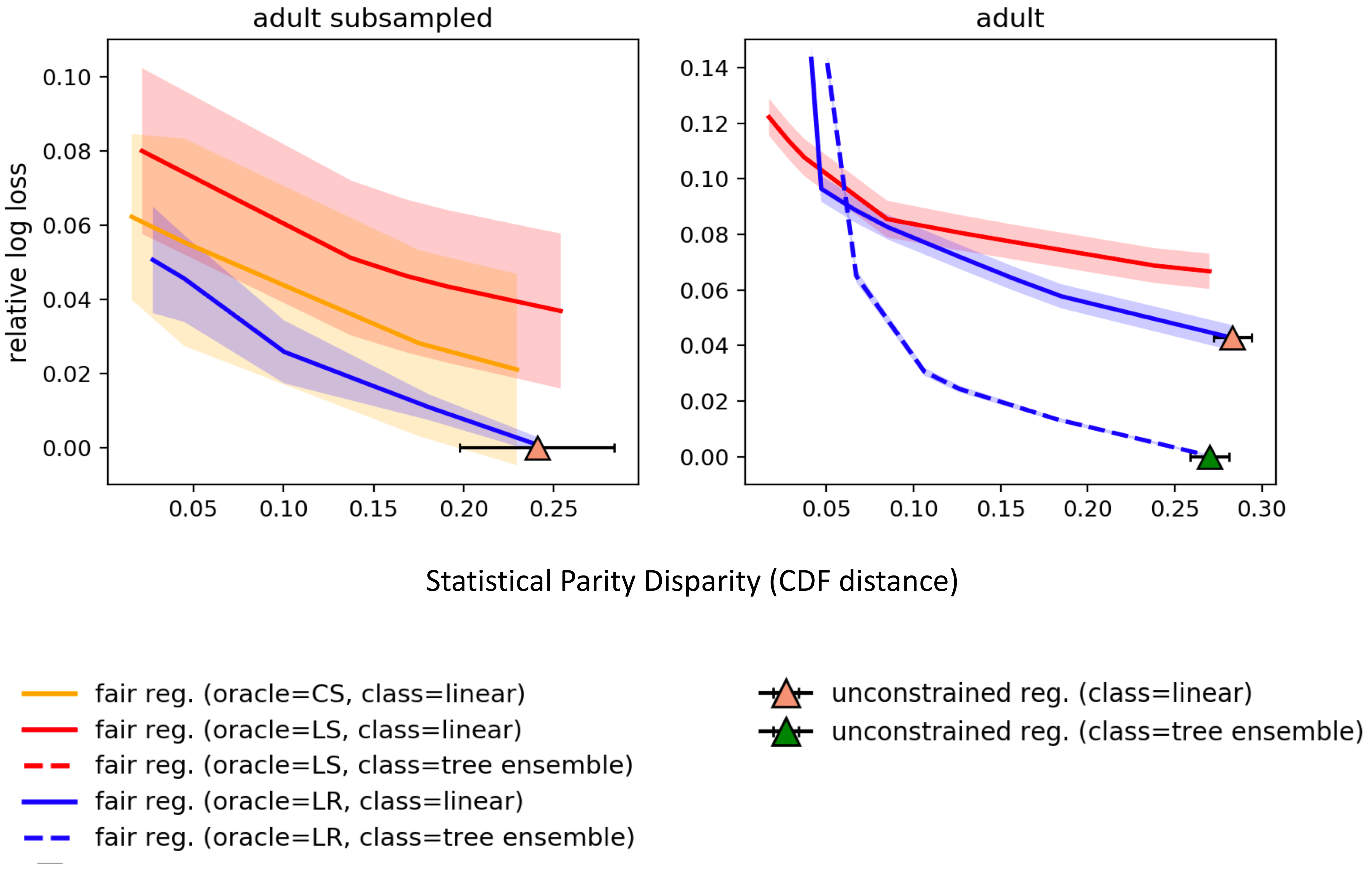
- Reduction-based algorithm: a provably efficient algorithms that iteratively solves a sequence of supervised learning problems (without fairness constraints):
  - Risk minimization under ℓ
  - Square loss minimization
  - Cost-sensitive classification (or weighted classification problem)

- Finite sample guarantees on:
  - Accuracy
  - Fairness violations

Empirical Evaluation

- Fairness constraint: statistical parity
- Data sets: Adult, Law School, Communities & Crime
- Losses: square loss, logistic loss
- Reductions:
  - Cost-sensitive classification (CS)
  - Square loss minimization (LS)
  - Logistic loss minimization (LR)
- Predictor classes: linear and tree ensemble


communities & crime
law school subsampled
law school
relative RMSE
Statistical Parity Disparity (CDF distance)
fair reg. (oracle=CS, class=linear)
fair reg. (oracle=LS, class=linear)
fair reg. (oracle=LS, class=tree ensemble)
SEO (class=linear)
unconstrained reg. (class=linear)
unconstrained reg. (class=tree ensemble)


adult subsampled
adult
relative log loss
0.10
0.08
0.06
0.04
0.02
0.00
0.14
0.12
0.05
0.15
0.20
0.25
0.30
Statistical Parity Disparity (CDF distance)
fair reg. (oracle=CS, class=linear)
fair reg. (oracle=LS, class=linear)
fair reg. (oracle=LS, class=tree ensemble)
fair reg. (oracle=LR, class=linear)
fair reg. (oracle=LR, class=tree ensemble)
unconstrained reg. (class=linear)
unconstrained reg. (class=tree ensemble)

Fair Regression:

Quantitative Definitions and Reduction-Based Algorithms

Poster: Thurs @ Pacific Ballroom #132