Algorithmic fairness in insurance pricing: A multi-class problem perspective

Insurance Data Science Conference,
17 - 18 June 2024

François HU - Head of Milliman R&D AI Lab

Joint work with Christophe Denis (LPSM, France), Mohamed Hebiri (LAMA, France) and Romuald Elie (LAMA, France)

Milliman

Summary

Algorithmic Fairness with multi-class problem perspective

1) Problem formulation

2) Approximate fairness

3) Some results in Insurance

Problem formulation

Fairness in multi-class problem with demographic parity

Milliman

Multi-class classification problems

Increasingly prevalent in actuarial studies

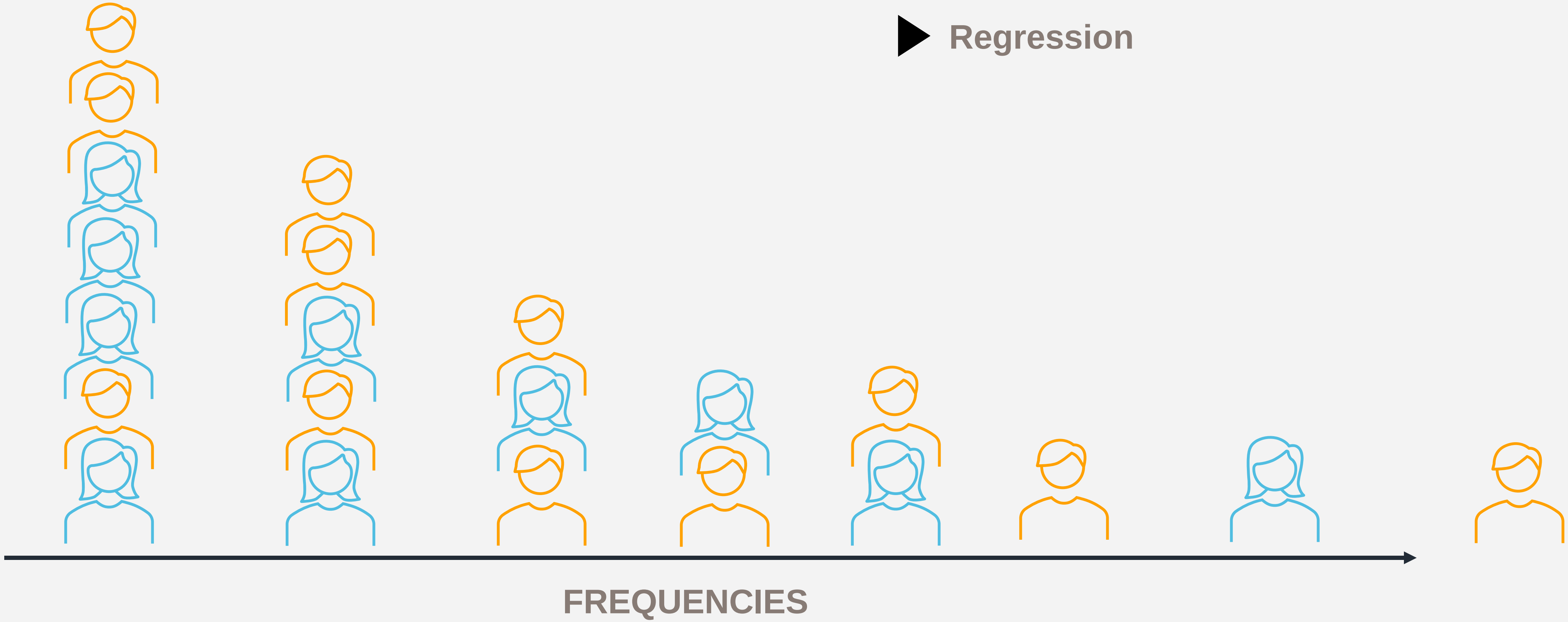
Illustrative example

Multi-class classification problems

Increasingly prevalent in actuarial studies

Illustrative example

▶ Regression



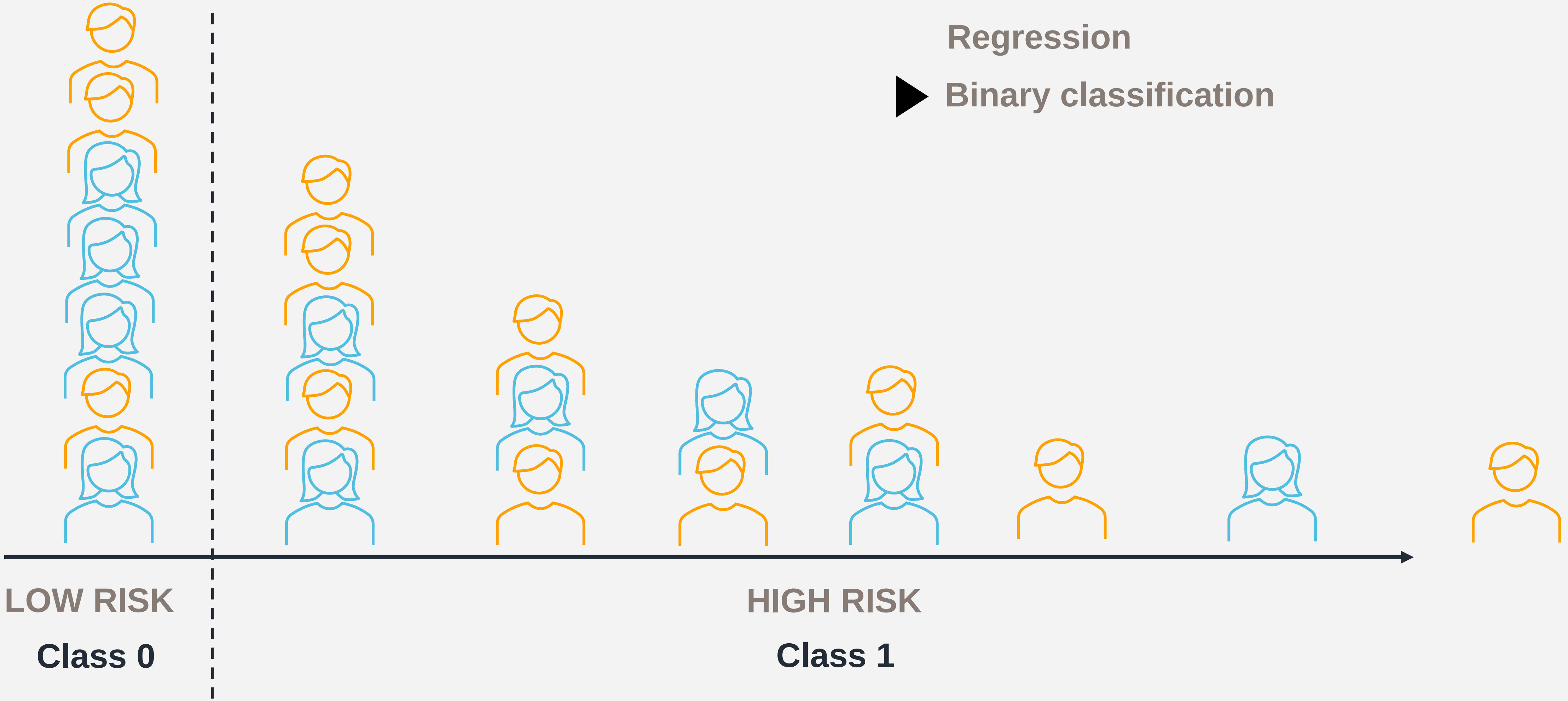
Multi-class classification problems

Increasingly prevalent in actuarial studies

Illustrative example

Regression

▶ Binary classification

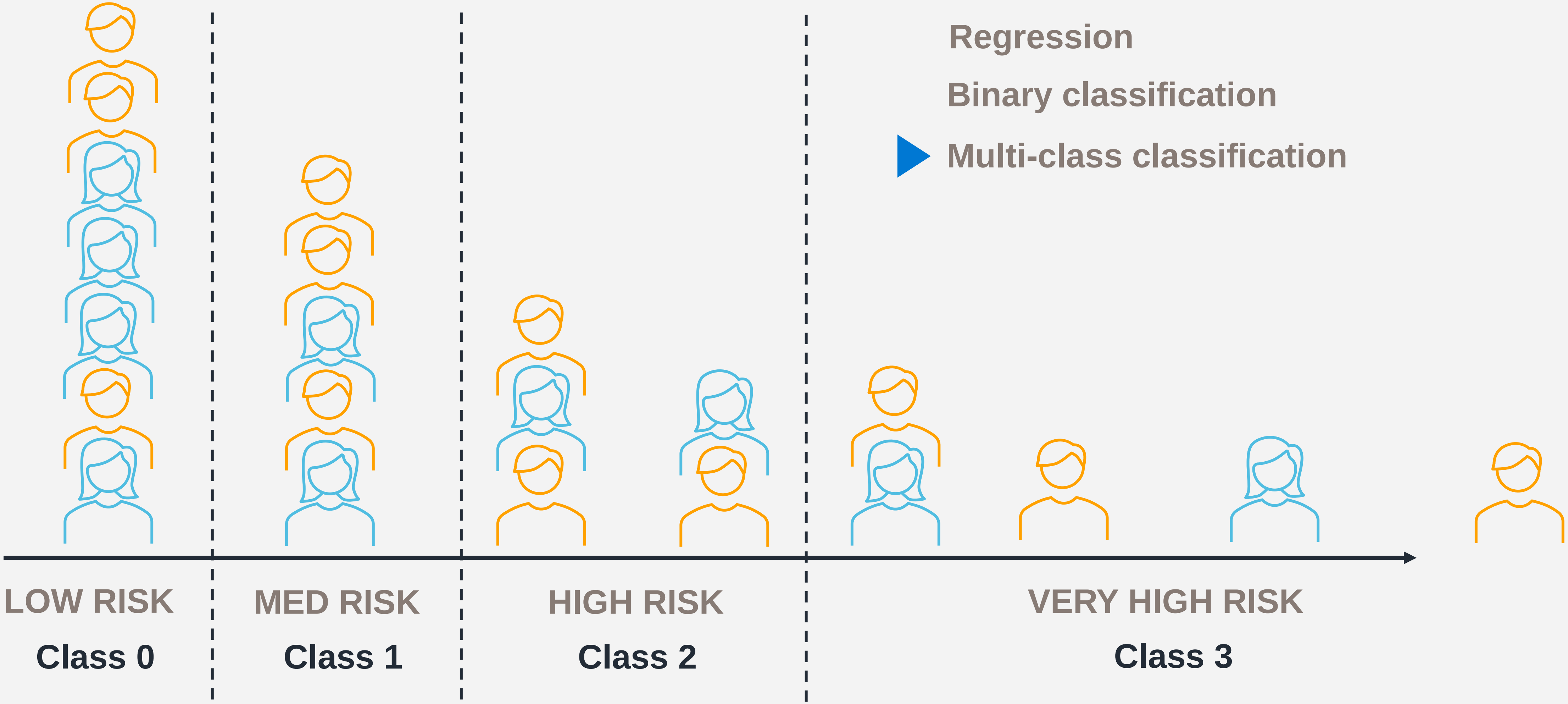


Multi-class classification problems

Increasingly prevalent in actuarial studies

Illustrative example

- Regression
- Binary classification
- ▶ Multi-class classification



Fairness in multi-class classification problem

Demographic Parity measure of fairness

Observations: $\Big(\underbrace{\text{features}}_{X}, \underbrace{\text{sensitive attribute}}_{S}, \underbrace{\text{label(s)}}_{Y}\Big)$

(Misclassification) Risk: $\mathcal{R}(g) = \mathbb{P}(g(X, S) \neq Y)$

Scores: $p_k(X, S) = \mathbb{P}(Y = k \mid X, S)$

Bayes classifier: $g^* \in \arg\min_g \left\{\mathcal{R}(g)\right\}$ ▶ $g^*(x, s) \in \arg\max_k p_k(x, s)$

Fairness in multi-class classification problem

Demographic Parity measure of fairness

Observations: $\Big(\underbrace{\text{features}}_{X},\ \underbrace{\text{sensitive attribute}}_{S},\ \underbrace{\text{label(s)}}_{Y}\Big)$

(Misclassification) Risk: $\mathcal{R}(g) = \mathbb{P}(g(X, S) \neq Y)$

Scores: $p_k(X, S) = \mathbb{P}(Y = k \mid X, S)$

Bayes classifier: $g^* \in \arg\min_g \left\{\mathcal{R}(g)\right\}$ ▶ $g^*(x, s) \in \arg\max_k p_k(x, s)$

▶ **Objective:** $g^*_{\text{fair}} \in \arg\min_g \left\{\mathcal{R}(g) : g(X, S) \perp S\right\}$ ▶ **Optimal fair** predictor **?**

Fairness in multi-class classification problem

Demographic Parity measure of fairness

Observations: $\Big(\underbrace{\text{features}}_{X},\ \underbrace{\text{sensitive attribute}}_{S},\ \underbrace{\text{label(s)}}_{Y}\Big)$

(Misclassification) Risk: $\mathcal{R}(g) = \mathbb{P}(g(X,\, S) \neq Y)$

Scores: $p_k(X, S) \;=\; \mathbb{P}(Y = k \mid X, S)$

Bayes classifier: $g^* \in \arg\min_g \Big\{\mathcal{R}(g)\Big\}$ ▶ $g^*(x, s) \in \arg\max_k p_k(x, s)$

▶ **Objective:** $g^*_{\text{fair}} \in \arg\min_g \Big\{\mathcal{R}(g) : g(X, S) \perp S\Big\}$ ▶ **Optimal fair** predictor **?**

Unfairness measure: $\mathcal{U}(g) = \max_k \left| \mathbb{P}(g(X, S) = k \mid S = 1) - \mathbb{P}(g(X, S) = k \mid S = -1) \right|$

Exact and Approximate Fairness

Optimal fair predictor and statistical guarantees

Milliman

Exact and approximate fairness

Risk: $\mathcal{R}(g) = \mathbb{P}(g(X, S) \neq Y)$

Unfairness measure: $\mathcal{U}(g) = \max_k \left| \mathbb{P}(g(X,S) = k \mid S = 1) - \mathbb{P}(g(X,S) = k \mid S = -1) \right|$

Exact fairness

$\mathcal{U}(g) = 0$

Approximate (or ε) fairness

$\mathcal{U}(g) \leq \varepsilon$

Milliman

Method of Langrange multipliers

Risk: $\mathcal{R}(g) = \mathbb{P}(g(X, S) \neq Y)$

Unfairness measure: $\mathcal{U}(g) = \max_k |\, \mathbb{P}(g(X,S) = k \mid S = 1) - \mathbb{P}(g(X,S) = k \mid S = -1)\,|$

Exact fairness

$\mathcal{U}(g) = 0$

Approximate (or ε) fairness

$\mathcal{U}(g) \leq \varepsilon$

Fair-risk

Lagrangian of the problem

$$\begin{aligned}\mathcal{R}_{\lambda^{(1)},\lambda^{(2)}}(g) := \mathcal{R}(g) &+ \sum_{k=1}^{K} \lambda_k^{(1)} [\mathbb{P}\left(g(X,S) = k|S = 1\right) - \mathbb{P}\left(g(X,S) = k|S = -1\right) - \varepsilon] \\ &+ \sum_{k=1}^{K} \lambda_k^{(2)} [\mathbb{P}\left(g(X,S) = k|S = -1\right) - \mathbb{P}\left(g(X,S) = k|S = 1\right) - \varepsilon]\end{aligned}$$

Optimal fair prediction and statistical guarantees (*)

Theorem (informal)

If $$(\lambda^{*(1)}, \lambda^{*(2)}) \in \arg\min_{(\lambda^{(1)},\lambda^{(2)})\in\mathbb{R}_+^{2K}} \sum_{s\in\mathcal{S}} \mathbb{E}_{X|S=s}\left[\max_k \left(\pi_s p_k(X,s) - s(\lambda_k^{(1)} - \lambda_k^{(2)})\right)\right] + \varepsilon \sum_{k=1}^{K} (\lambda_k^{(1)} + \lambda_k^{(2)})$$

then $$g^*_{\varepsilon-\text{fair}}(x,s) = \arg\max_{k\in[K]} \left(\pi_s p_k(x,s) - s(\lambda_k^{*(1)} - \lambda_k^{*(2)})\right)$$

- **Closed-form** solution
- **Post-processing** and **model agnostic**
- **Theorem (informal) :** a plug-in estimator makes the model asymptotically as performant as $g^*_{\varepsilon\text{-fair}}$ in terms of **fairness** and **predictive performance**

(*) post-processing approach for multi-classes https://github.com/HsiangHsu/Fair-Projection (NeurIPS 2022)

Numerical evaluation (*)

Dataset: DRUG, CRIME

Machine Learning Models:

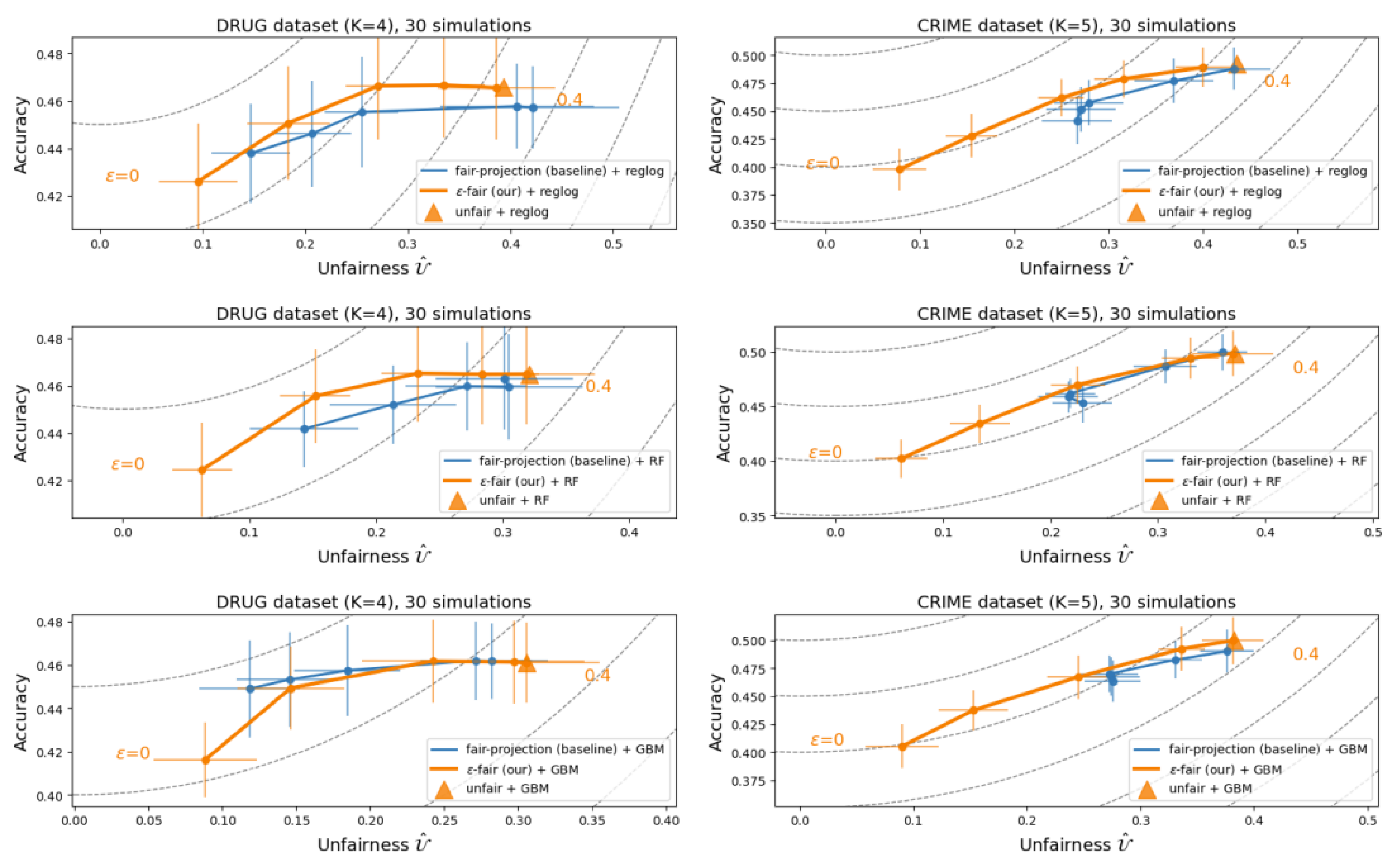
Logistic regression (**GLM**)

Random Forest (**RF**)

LightGBM (**GBM**)

Benchmark:

Fair-projection (1)



(*) **Our paper**: fairness guarantees in multi-class classification with demographic parity (JMLR 2024)

(1) **Baseline**: post-processing approach for multi-classes (NeurIPS 2022)

Insurance dataset

Car insurance portfolio

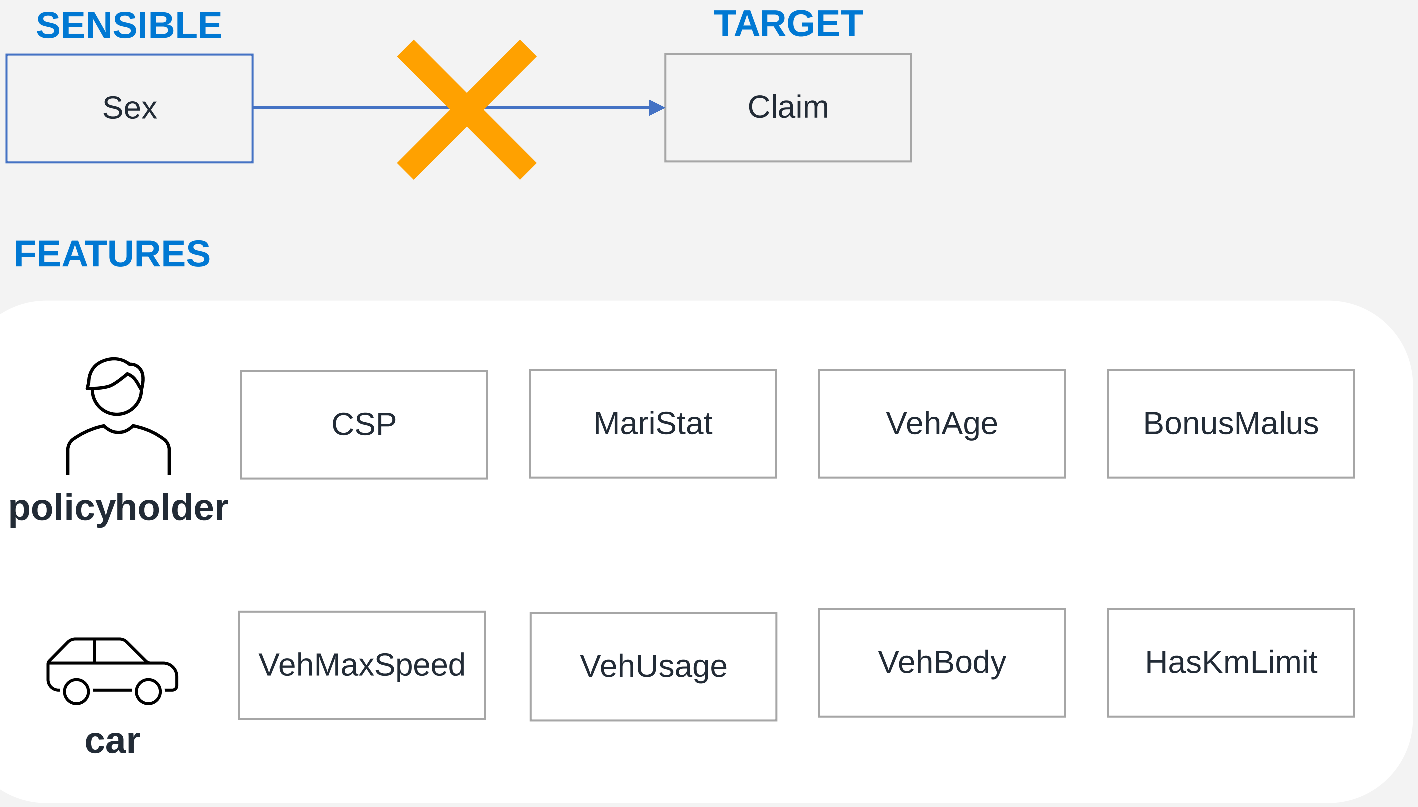
Milliman

Description of the dataset

Toy example: automobile insurance portfolio

Real-world example

The **freMPL** dataset (CASdataset) is a database used in the automobile insurance industry. It contains information on driver characteristics, insured vehicles and associated claims (+10k observations).

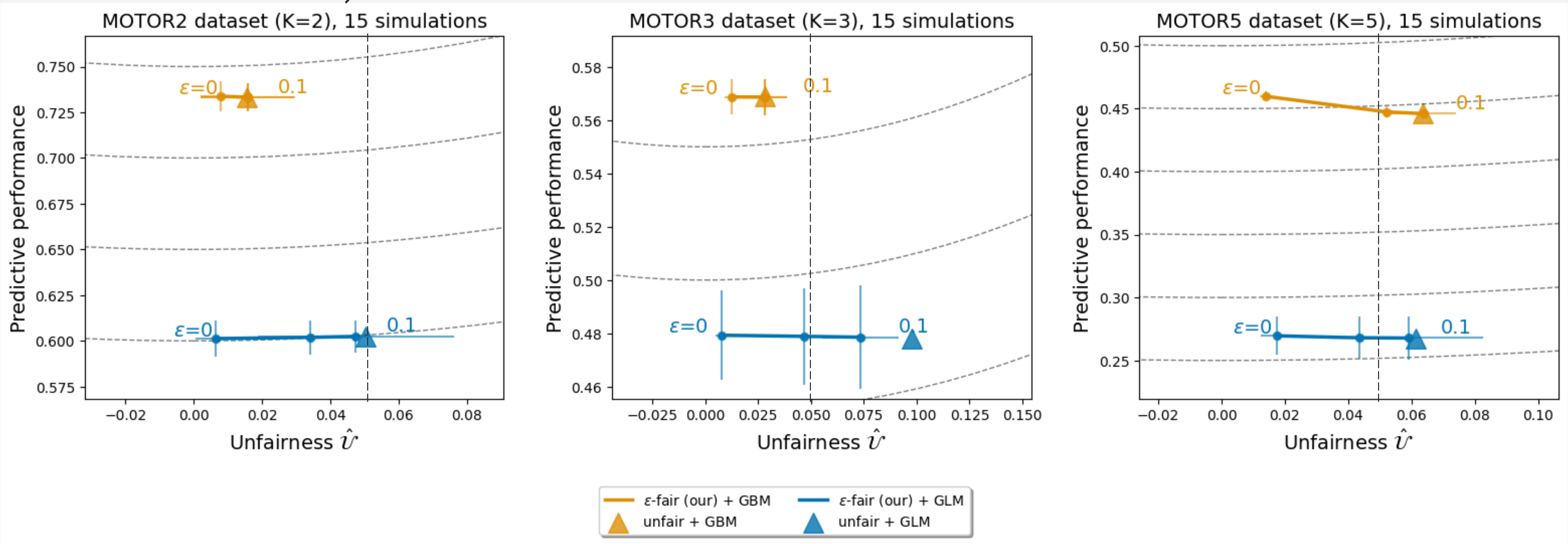


Numerical evaluation (1/2)

Toy example: automobile insurance portfolio

Base GBM seems **DP-fair**

Base GBM seems **DP-unfair**

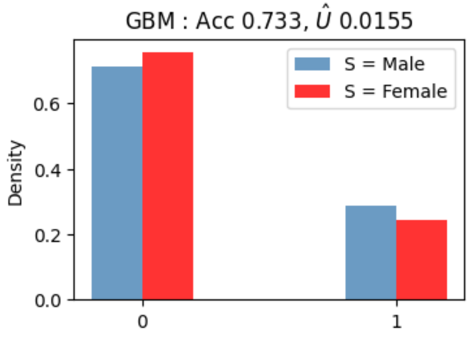


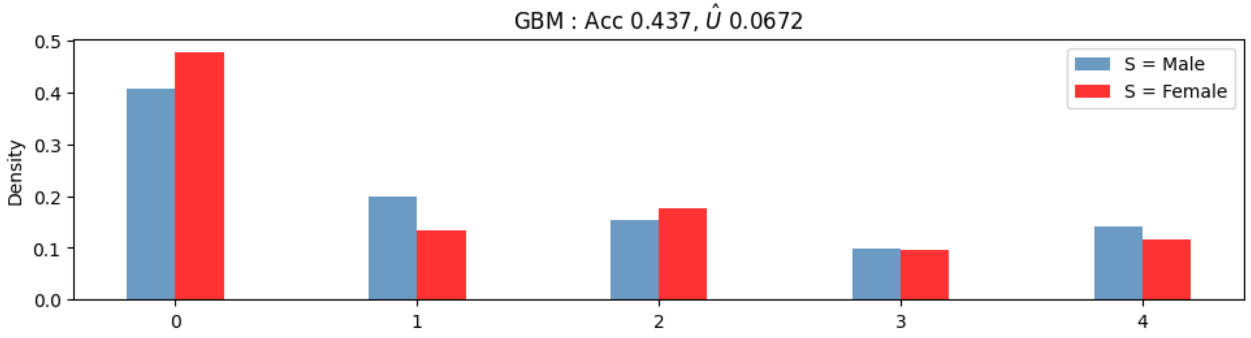
Numerical evaluation (2/2)

Toy example: automobile insurance portfolio

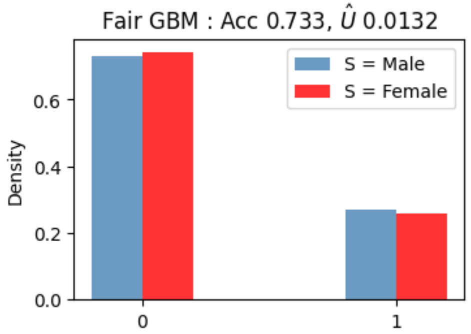
2 classes | 5 classes

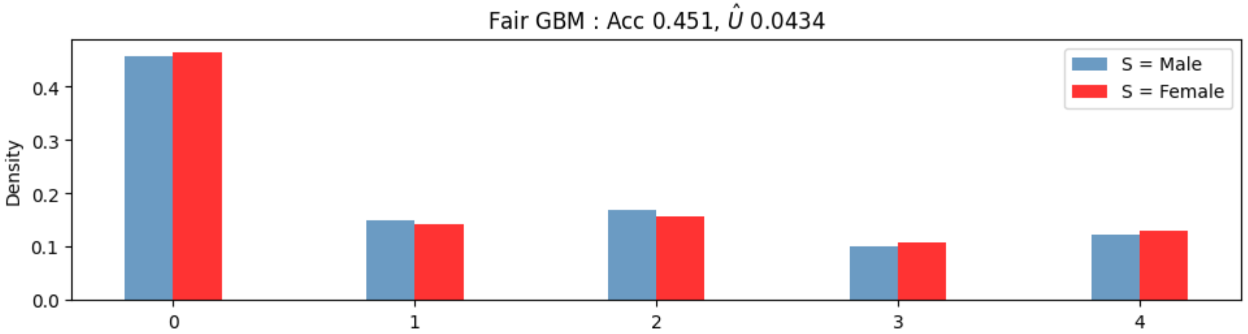
Before remediation





After remediation





In summary

- Multi-class classification paradigm enables precise risk categorization, enhancing accuracy in insurance pricing.
- Post-processing approach applicable to any off-the-shelf ML model
- Compared to regression tasks, multi-class framework can better achieve other fairness metrics like separation and sufficiency.

Thank you for your attention !