Classification and Regression under Statistical Parity

Solenne Gaucher [1]

Evgenii Chzhen [2]

Vincent Divol[3]

Nicolas Schreuder [4]

[1]CREST ENSAE, IPP

[2]CNRS, LMO, Université Paris-Saclay

[3]CEREMADE, Université Paris Dauphine PSL

[4]CNRS, LIGM, Université Gustave Eiffel

Table of contents

~~Introduction to statistical fairness~~

Setting

Observations: $(\underbrace{X}_{\text{feature}}, \underbrace{S}_{\text{sensitive attribute}}, \underbrace{Y}_{\text{outcome}}) \sim \mathbb{P}$ on $\mathcal{X} \times \mathcal{S} \times \mathcal{Y}$.

Setting	Regression
Outcome	$\mathcal{Y} = \mathbb{R}$
Predictions	$\mathcal{F} \coloneqq \{f : \mathcal{X} \times \mathcal{S} \to \mathbb{R}\}$
Risk	$\mathcal{R}(f) = \mathbb{E}\,(Y - f(X, S))^2$
Fairness criteria	$f(X, S) \perp\!\!\!\perp S$ (Statistical Parity)

Setting

Observations: $(\underbrace{X}_{\text{feature}}, \underbrace{S}_{\text{sensitive attribute}}, \underbrace{Y}_{\text{outcome}}) \sim \mathbb{P}$ on $\mathcal{X} \times \mathcal{S} \times \mathcal{Y}$.

Setting	Regression	Classification
Outcome	$\mathcal{Y} = \mathbb{R}$	$\mathcal{Y} = \{0, 1\}$
Predictions	$\mathcal{F} := \{f : \mathcal{X} \times \mathcal{S} \to \mathbb{R}\}$	$\mathcal{G} := \{g : \mathcal{X} \times \mathcal{S} \to \{0, 1\}\}$
Risk	$\mathcal{R}^{\mathrm{mse}} = \mathbb{E}\,(Y - f(X, S))^2$	$\mathcal{R}^{0-1} = \mathbb{E}\,[Y \neq g(X, S)]$
Fairness criteria	$f(X, S) \perp\!\!\!\perp S$ (Statistical Parity)	$\mathbb{E}[g(X, S) \mid S] \perp\!\!\!\perp S$ (Demographic Parity)

Setting

Observations: $(\underbrace{X}_{\text{feature}}, \underbrace{S}_{\text{sensitive attribute}}, \underbrace{Y}_{\text{outcome}}) \sim \mathbb{P}$ on $\mathcal{X} \times \mathcal{S} \times \mathcal{Y}$.

Setting	Regression	Classification
Outcome	$\mathcal{Y} = \mathbb{R}$	$\mathcal{Y} = \{0, 1\}$
Predictions	$\mathcal{F} := \{f : \mathcal{X} \times \mathcal{S} \to \mathbb{R}\}$	$\mathcal{G} := \{g : \mathcal{X} \times \mathcal{S} \to \{0, 1\}\}$
Risk	$\mathcal{R}(f) = \mathbb{E}(Y - f(X, S))^2$	$\mathcal{R}^{0-1}(g) = \mathbb{E}[Y \neq g(X, S)]$
Fairness criteria	$f(X, S) \perp\!\!\!\perp S$ (Statistical Parity)	$\mathbb{E}[g(X, S) \mid S] \perp\!\!\!\perp S$ (Demographic Parity)

This is the awareness framework... in the unawareness framework, f and g cannot depend on S.

Observations: $(\underbrace{t}_{\text{feature}}; \underbrace{\hat{h}}_{\text{sensitive attribute}}; \underbrace{\tilde{U}}_{\text{outcome}})$ P on X S Y.

Setting	Regression	Classification
Outcome	$Y = \mathrm{R}$	$Y = f0; 1g$
Predictions	$F\ ,\ f\grave{\imath} : Z\ !\ \mathrm{R}g$	$G\ ,\ f\hat{O} : Z\ !\ f0; 1gg$
Risk	$R^{\text{'tc}} = \mathrm{E}\ (\tilde{U}\ \ \grave{\imath}(\hat{W}))^2$	$R^{0\ \ 1} = \mathrm{E}\,[\tilde{U} \neq \hat{O}(\hat{W})]$
Fairness criteria	$\grave{\imath}(\hat{W})\ ??\ \hat{h}$ (Statistical Parity)	$\mathrm{E}[\hat{O}(\hat{W}) j \hat{h}]\ ??\ \hat{h}$ (Demographic Parity)

Awareness framework : $Z = X\ \ S$, $\hat{W} = (t; \hat{h})$

Unawareness framework : $Z = X$, $\hat{W} = t$

Without fairness constraint, the Bayes-optimal solution to

$$\underset{g \in \mathcal{G}}{\text{minimize}} \quad \mathbb{P}[Y \neq g(X)]$$

is given by

$$g^{\text{Bayes}}(x) = \mathbb{1}\{\eta(x) \geq \tfrac{1}{2}\}$$

where $\eta(X) := \mathbb{E}[Y|X]$ is the solution to

$$\underset{f \in \mathcal{F}}{\text{minimize}} \quad \mathbb{E}\,(Y - f(X))^2 \ .$$

This relationship between classification and regression can be used to design and study classifiers [Yang, 1999], [Massart and Nédélec, 2006], [Audibert and Tsybakov, 2007], [Biau et al., 2008]

Reminders on unconstrained classification

<u>Consider the risks</u>

$$R \ (\hat{O}) = \ \mathrm{P}\,[\tilde{U} = 0; \hat{O}(\hat{W}) = 1] + (1 \qquad)\mathrm{P}\,[\tilde{U} = 1; \hat{O}(\hat{W}) = 0] :$$

We have

$$R \ (\hat{O}) = (1 \qquad)\mathrm{E}[\tilde{U}] + \mathrm{E}\,[\hat{O}(\hat{W})(\qquad (\hat{W}))] :$$

$=)$ The Bayes classifier $\hat{O}^{@h\grave{\imath}\, \bar{\dot{a}}\,'}$ is given by

$$\hat{O}^{@h\grave{\imath}\, \bar{\dot{a}}\,'}(\_\,) = 1\, f \ (\_\,) \qquad g :$$

There is an equivalence between solving the regression problem and solving the classification problem for all .

What about fair classification?

Let $\grave{\imath}$ be the solution to

$$\text{minimize} \quad \mathrm{E}\ (\tilde{U} \quad \grave{\imath}(\hat{W}))^2$$

$$\text{such that} \quad \grave{\imath}(\hat{W}) \ ?? \ \hat{h}:$$

We can define $\hat{O} :\_ \ 7! \ 1 f \grave{\imath} \ (\_) \qquad g.$

- Is $\hat{O}$ optimal for some threshold ?
- Does $\hat{O}$ verify Demographic Parity?

What about fair classification?

Let $\grave{\imath}$ be the solution to

$$\begin{aligned} \text{minimize} \quad & \mathrm{E}\ (\tilde{U} \quad \grave{\imath}(\hat{W}))^2 \\ \text{such that} \quad & \grave{\imath}(\hat{W}) \ ?? \ \hat{h}: \end{aligned}$$

We can define $\hat{O} :\_ \ 7! \ 1 \ f\grave{\imath} \ (\_) \qquad g.$

- Is $\hat{O}$ optimal for some threshold ?
- Does $\hat{O}$ verify Demographic Parity? Yes… it verifies **Strong Demographic Parity**

What about fair classification?

Let $\grave{\imath}$ be the solution to

$$\text{minimize} \quad \mathrm{E}\ (\tilde{U} \quad \grave{\imath}(\hat{W}))^2$$
$$\text{such that} \quad \grave{\imath}(\hat{W}) \;⁇\; \hat{h}:$$

We can define $\hat{O} :\_ \;7! \;1\, f\grave{\imath}\; (\_\,) \qquad g.$

- Is $\hat{O}$ optimal for some threshold ?
- Does $\hat{O}$ verify Demographic Parity? Yes… it verifies **Strong Demographic Parity**

Strong Demographic parity [JPSJC19]: A classifier $\hat{O}$ verifies Strong Demographic parity if $\hat{O}(\_\,) = 1\, f\grave{\imath}(\_\,) \qquad g$ for some threshold , and $\grave{\imath}$ verifies Statistical Parity.

What about fair classification?

Let $\grave{\imath}$ be the solution to

$$\begin{array}{ll} \text{minimize} & \mathrm{E}\ (\tilde{U} \quad \grave{\imath}(\hat{W}))^2 \\ \text{such that} & \grave{\imath}(\hat{W}) \ ?? \ \hat{h}: \end{array}$$

We can define $\hat{O} :\_ \ 7! \ 1\ f\grave{\imath}\ (\_ \ ) \qquad g.$

- Is $\hat{O}$ optimal for some threshold ? It depends...
- Does $\hat{O}$ verify Demographic Parity? Yes... it verifies **Strong Demographic Parity**

Strong Demographic parity [JPSJC19]: A classifier $\hat{O}$ verifies Strong Demographic parity if $\hat{O}(\_ \ ) = 1\ f\grave{\imath}(\_ \ ) \qquad g$ for some threshold , and $\grave{\imath}$ verifies Statistical Parity.

The awareness framework

Fair regression

Reminders from Nicolas's talk

Assumption (A1)

[illegible] $(x; \hat{h}) | \hat{h} =$ [illegible]

Theorem (Chzhen et al., 2020, Le Gouic et al., 2020)

[illegible]

$$[\text{illegible}] \quad \mathbb{E}\left[(\tilde{U} \quad [\text{illegible}](x; \hat{h}))^2 \right]$$

$$[\text{illegible}] \quad [\text{illegible}](x; \hat{h}) \quad [\text{illegible}] \quad \hat{h}.$$

[illegible]

$$[\text{illegible}] \; (;) = [\text{illegible}]^{-1} \quad [\text{illegible}] \; (\; (;))$$

[illegible] $(\,) = \mathrm{P}[\;(x; \hat{h}) \quad | \hat{h} =$ [illegible] $]$ [illegible] $^{-1}(\;) = \sum$ [illegible] $^{1}(\;)$ [illegible]

Fair classification

First results

Assumption (A2)

$\cdot$ ꟿ $Íh\!\!\!\!\text{ʊ}''$ ($(t;\hat{h})j\hat{h} =$ ’) ƕʬā !ωɧ¯ɧ˒ω˒ ’ ƕɧă ’˒ ʢʢωʬ¯āă ωɧ ƕɧ ɧ¯āʬ˙ƕɏ

Theorem (G., Schreuder and Chzhen, 2023)

Íā¯ Ô ɛā ¯ȡā ’ωɏ˒ ¯ωɧ ¯ω

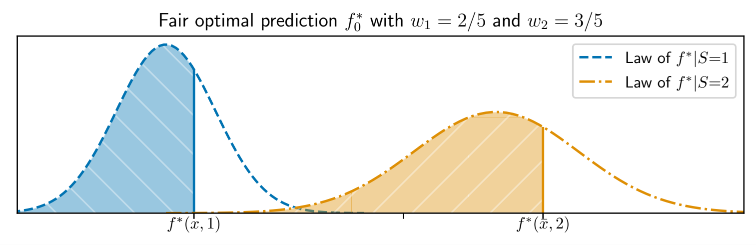
$$g\,ɧ\,g\,\_\,ā \qquad \mathrm{P}\,[\tilde{U} \in \hat{O}(t;\hat{h})]$$

$$’˒\,!ȡ\,¯ȡƕ¯ \qquad \hat{O}(t;\hat{h}) \;?? \;\hat{h}:$$

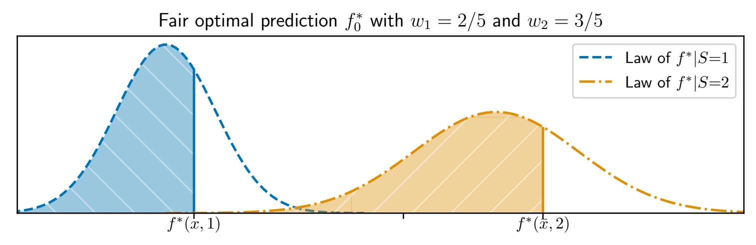
Ŀɧăāʬ ’’˒ g ʢ¯ωɧ ʖ ╖ Ô !ƕɧ ɛā ā’ ʢʬā’’āă ƕ’

$$\hat{O}\;(’;’) = 1 \quad \grave{\imath}\;(’;’) \quad \frac{1}{2} \quad :$$

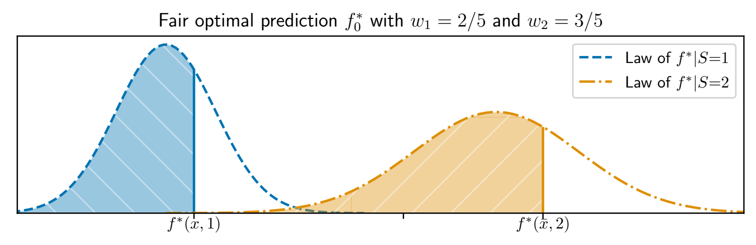
We look for classifiers Ô(’ ; ’) = 1 f (’ ; ’) ›g :



We look for classifiers Ô(’ ; ’) = 1 f (’ ; ’) ,g :



We look for classifiers Ô(’ ; ’) = 1 f (’ ; ’) ›g :



We consider general performance measures

$$U_{n;d}(\hat{O}) \; , \; \frac{n_0 + n_1 P[\hat{O}(t;\hat{h}) = 1; \tilde{U} = 1] + n_2 P[\hat{O}(t;\hat{h}) = 1]}{d_0 + d_1 P[\hat{O}(t;\hat{h}) = 1; \tilde{U} = 1] + d_2 P[\hat{O}(t;\hat{h}) = 1]}$$

where n, d can depend on $P[\tilde{U} = 1]$ (but not on $\hat{O}$).

- I Accuracy $U_{n;d}(\hat{O}) = P[\hat{O}(t;\hat{h}) \neq \tilde{U}]:$
- I $\{_\varepsilon$-score $U_{n;d}(\hat{O}) = \frac{(1+\varepsilon^2) P[\hat{O}(t;\hat{h}) = 1; \tilde{U} = 1]}{\varepsilon^2 P[\tilde{U} = 1] + P[\hat{O}(t;\hat{h}) = 1]}:$
- I Jaccard Index $U_{n;d}(\hat{O}) = \frac{P[\hat{O}(t;\hat{h}) = 1; \tilde{U} = 1]}{P[\hat{O}(t;\hat{h}) = 1; \tilde{U} = 0] + P[\hat{O}(t;\hat{h}) = 1]}:$
- I AM Measure $U_{n;d}(\hat{O}) = \frac{1}{2}(P[\hat{O}(t;\hat{h}) = 1 j \tilde{U} = 1] + P[\hat{O}(t;\hat{h}) = 0 j \tilde{U} = 0]):$

Results

Theorem (G., Schreuder and Chzhen, 2023 - Informal)

[illegible] $\hat{O}_{n;d}$ [illegible]

$$[\text{illegible}] \qquad U_{n;d}(\hat{O})$$

$$[\text{illegible}] \qquad \hat{O}(t;\hat{h}) \;[\text{illegible}]\; \hat{h}.$$

[illegible] $\hat{O}_{n;d}$ [illegible]

$$\hat{O}_{n;d}(\cdot;\cdot) = 1 \;[\text{illegible}]\; (\cdot;\cdot) \quad [\text{illegible}]_{n;d} \; .$$

Assumption (A3)

1. $d_0 + (d_1 + (d_2)_+)_+ \quad 0$
2. $\begin{cases} d_2 n_1 > n_2 d_1 \quad d_0 n_1 \quad n_0 d_1 \quad (n_0 d_2 \quad d_0 n_2)_+ \\ \frac{n_0 d_2 \quad d_0 n_2}{n_2 d_1 \quad d_2 n_1} \quad P[\tilde{U} = 1] \end{cases}$
3. $\begin{cases} d_2 n_1 = n_2 d_1 \quad n_1 d_0 > d_1 n_0 \\ \frac{d_0 n_2 \quad n_0 d_2}{n_0 d_1 \quad d_0 n_1} \; 2 \; [0;1] \end{cases}$

Theorem (G., Schreuder and Chzhen, 2023)

$\hat{O}_{n;d}(\cdot;\cdot) = 1 \quad (\cdot;\cdot) \quad {}_{n;d}$

- ${}_{n;d}$

$$E[(\; (t;\hat{h}) \quad)_+] = \frac{n_0 d_1 \quad d_0 n_1}{n_2 d_1 \quad d_2 n_1} + \frac{n_0 d_2 \quad d_0 n_2}{n_2 d_1 \quad d_2 n_1}$$

- ${}_{(n;d)} = \frac{d_0 n_2 \quad n_0 d_2}{n_0 d_1 \quad d_0 n_1};$

Algorithm:

1. Estimate
2. Estimate ì
3. Estimate n and d
4. Estimate threshold $_{n;d}$ (explicit formula or fixed-point equation)
5. Use double-plug-in estimator

$$\hat{b}_{n;d}(' ; ') = 1 \left\{ \hat{\imath}(' ; ') \quad \hat{b}_{n;d} \right\}$$

Summary

In the awareness framework, DP-fair optimal classification with performance measure $U_{n;d}$

- is given by $\hat{O}_{n;d}(' \,;\, ') = 1 \left\{ ì\,(' \,;\, ') \quad {}_{n;d} \right\}$
- has desirable properties :
  - **does no harm to the protected group**
  - preserves **rational ordering**
  - preserves **monotonicity**

Summary

In the awareness framework, DP-fair optimal classification with performance measure $U_{n;d}$

- is given by $\hat{O}_{n;d}(' ; ') = 1 \; n \; \grave{\hat{\imath}} \; (' ; ') \quad {}_{n;d} \; o$
- has desirable properties :
  - **does no harm to the protected group**
  - preserves **rational ordering**
  - preserves **monotonicity**

(Informal) A family of classifiers $\hat{O}$ preserves **monotonicity** if -almost-surely,

$$f P[\hat{O}\ (t; \hat{h}) = 1] > P[\hat{O}_{\ 0}(t; \hat{h}) = 1] g \quad =) \quad f\hat{O}\ (' ; ') \quad \hat{O}_{\ 0}(' ; ') g.$$

The unawareness framework

Fair regression

Fair regression under unawareness

Problem: We want to solve

$$\text{minimize} \quad \mathrm{E} \; (\tilde{U} \quad \breve{\imath}(t))^2$$
$$\text{such that} \quad \breve{\imath}(t) \; ?? \; \hat{h}:$$

In the following, we assume $S = f1; 2g$.

Notations: $t \quad , \quad tj\hat{h} = 1 \quad _1, \quad tj\hat{h} = 2 \quad _2.$

An equivalent fairness constraint

Jordan decomposition We write ${}_1$ ${}_2 =$ $({}_+$), where

- ${}_+$ and are positive measures
- X_+ , supp(${}_+$) and X , supp( ) are disjoint.

An equivalent fairness constraint

Jordan decomposition We write $_1$ $_2$ = *Ù*($_+$), where

- $_+$ and are probability measures
- X_+ , supp($_+$) and X , supp( ) are disjoint.

An equivalent fairness constraint

Jordan decomposition We write $\ _1 \quad _2 = \dot{U}(\ _+ \quad)$, where

- $_+$ and $\quad$ are probability measures
- $X_+ \ ,$ supp($_+$) and $X \ ,$ supp($\ $) are disjoint.

Lemma (Chzhen and Schreuder, 2020)

$: X \to \mathbb{R}$ [illegible]

$$]_{+} =] \quad :$$

An equivalent fairness constraint

Jordan decomposition We write $_1 \quad _2 = Ù(\; _+ \quad)$, where

- $_+$ and are probability measures
- $X_+ \; ,$ supp($_+$) and $X \; ,$ supp( ) are disjoint.

Lemma (Chzhen and Schreuder, 2020)

[illegible]

$$]\; _+ =]$$

Consequence We can look for functions

$$(\,) = \begin{cases} _+(\,) & \text{if} \; 2 \; X_+ \\ (\,) & \text{if} \; 2 \; X \\ (\,) & \text{if} \; 2 \; X_= \; , \; f \; : \; _1(\,) = \; _2(\,) g : \end{cases}$$

such that $_+ \,] \; _+ = \;]$.

$$R^{'tc}(ì) = \int_{X_+} ! \left((') ; (\; ') \; ; ì_+ (') \right) ǎ_{+} (') + \int_{X} ! \left((') ; (\; ') \; ; ì \; (') \right) ǎ \; (') + ! \, ' \bar{ǎ}$$

where $! : \mathrm{R} \quad \mathrm{R}_+ \quad \mathrm{R} \; ! \; \mathrm{R}_+$ is a cost, and $(') = \begin{cases} \frac{ǎ_{+}}{ǎ} (') \text{ if } ' \; 2 \; X_+ \\ \frac{ǎ}{ǎ} (') \text{ if } ' \; 2 \; X \end{cases}$.

$$R^{'tc}(\grave{\imath}) = \underset{X_+}{Z} \; ! \; ('); \; (\;') \; ; \grave{\imath}_+ (') \;\; \check{a} \;\; _+ (') + \underset{X}{Z} \; ! \; ('); \; (\;') \; ; \grave{\imath} \;\; (') \;\; \check{a} \;\; (') + !'\bar{\check{a}}$$

where $! : \quad \mathbb{R} \quad \mathbb{R}_+ \qquad \mathbb{R} \;\; ! \;\; \mathbb{R}_+$ is a cost, and $(') = \begin{cases} \frac{\check{a} \;\; _+}{\check{a}} (') \text{ if } ' \;\; 2 \;\; X_+ \\ \frac{\check{a}}{\check{a}} (') \text{ if } ' \;\; 2 \;\; X \end{cases}$.

I $\grave{\imath}$ should depend only on $('); \;\; (') \;\; : \grave{\imath} \;\; (') = \tilde{\grave{\imath}} \;\; ('); \;\; (') \;\; .$

$$\mathrm{R}^{\,'tc}(\grave{\imath}) = \int_{X_+} ! \quad (')\,;\,(\quad ') \quad ;\grave{\imath}_+(') \quad \check{a} \quad {}_+(') + \int_{X} ! \quad (')\,;\,(\quad ') \quad ;\grave{\imath} \quad (') \quad \check{a} \quad (') + \; !\,'\,\bar{\check{a}}$$

where $! : \quad \mathrm{R} \quad \mathrm{R}_+ \qquad \mathrm{R} \; ! \; \mathrm{R}_+$ is a cost, and $(') = \begin{cases} \frac{\check{a} \quad +}{\check{a}}(') \text{ if } ' \; 2 \; X_+ \\ \frac{\check{a}}{\check{a}}(') \text{ if } ' \; 2 \; X \end{cases}$.

- $\grave{\imath}$ should depend only on $(')\,; \quad (') \; : \grave{\imath} \quad (') = \tilde{\imath} \qquad (')\,; \quad (')$.
- Let $\sim \; =$ Law $(\; (t)\,; \quad (t)) \quad t$. $\tilde{\imath}$ should solve:

$$\text{minimize} \int ! \; (\; ; \;) \; ;\tilde{\imath}_+(\; ; \;) \quad \check{a}\sim_+(\; ; \;) + \int ! \; (\; ; \;) \; ;\tilde{\imath} \; (\; ; \;) \quad \check{a}\sim \; (\; ; \;)$$

$$\text{such that} \quad \tilde{\imath}_+ \,]\sim_+ \; = \; \tilde{\imath} \quad]\sim$$

I Let ~ =Law ((t); (t)) t . ı̃ should solve:

minimize Z ! (;) ;ı̃+ (;) ă~+ (;) + Z ! (;) ;ı̃ (;) ă~ (;)

such that ı̃+] ~+ = ı̃] ~

- Let ~ = Law ((t); (t)) t . ĩ should solve:

$$\text{minimize} \int ! \; (\; ; \;) \; ; ĩ_+ (\; ; \;) \; ǎ \sim_+ (\; ; \;) + \int ! \; (\; ; \;) \; ; ĩ \; (\; ; \;) \; ǎ \sim \; (\; ; \;)$$

such that ĩ+] ~+ = ĩ] ~

- ĩ+] ~+ = ĩ] ~ should solve the barycenter problem:

$$\min \ddot{e} \dot{I}_! (\sim_+ ; \;) + \ddot{e} \dot{I}_! (\sim \; ; \;)$$

where

$$\ddot{e} \dot{I}_! (\sim ; \;) = \inf_{2 \, (\sim ; \;)} \int !((\; ; \;); ˛) ǎ \; ((\; ; \;); ˛):$$

Result

Assumption (A4)

[illegible]

Theorem (Divol and G., 2024)

[illegible]

$$\min [illegible](\sim_+ ; \ ) + [illegible](\sim \ ; \ )$$

[illegible]

$$[illegible](\cdot) = \begin{cases} [illegible]_+ (\ (\cdot); \ (\cdot)) & [illegible] \ \cdot \in X_+ \\ [illegible] (\ (\cdot); \ (\cdot)) & [illegible] \ \cdot \in X \\ (\cdot) & [illegible]. \end{cases}$$

Remarks:

- Under Assumptions (A1) and (A4), the optimal prediction is **deterministic**.
- The optimal prediction tries to guess the sensitive attribute.
- Unless verifies Statistical Parity, the optimal fair prediction does not verify **rational ordering** within group.

Fair classification

Fair classification under unawareness

We want to solve

$$\text{minimize} \quad R \ (\hat{O}) \ , \quad \mathrm{P}[\tilde{U} = 0; \hat{O}(ŧ) = 1] + (1 \qquad)\mathrm{P}[\tilde{U} = 1; \hat{O}(ŧ) = 0]$$

$$\text{such that} \quad \hat{O}(ŧ) \ ?? \ \hat{h}:$$

Remark:

$$R \ (\hat{O}) = (1 \qquad)\mathrm{E}[\tilde{U}] + \mathrm{E}\,[\hat{O}(ŧ)(\qquad (ŧ))]:$$

$=)$ The Bayes classifier $\hat{O}^{@\hbar\upsilon` \bar{\check{a}}'}$ is given by

$$\hat{O}^{@\hbar\upsilon` \bar{\check{a}}'}(') = 1 \ f \ (') \qquad g:$$

I Family of risks R with corresponding thresholds .

Lemma (Chzhen and Schreuder, 2020)

$$: X \to \mathbb{R}$$

$$] =] :$$

- $\hat{O}$ verifies Demographic Parity if and only if

$$\mathbb{E}_{+}[\hat{O}(t)] = \mathbb{E}[\hat{O}(t)]:$$

Lemma (Chzhen and Schreuder, 2020)

$: X \to \mathbb{R}$ [illegible]

$$] _{+} =] \quad :$$

- $\hat{O}$ verifies Demographic Parity if and only if

$$\mathbb{E}_{+}[\hat{O}(t)] = \mathbb{E}[\hat{O}(t)]:$$

- We can look for $\hat{O}(') = \begin{cases} \hat{O}_{+}(') & \text{if } ' \in X_{+} \\ \hat{O}(') & \text{if } ' \in X \\ 1\{f(') \quad g\} & \text{else:} \end{cases}$

Lemma (Chzhen and Schreuder, 2020)

$: X \; ! \; \mathrm{R}$

$$]_{+} =] \; :$$

- $\hat{O}$ verifies Demographic Parity if and only if

$$\mathrm{E}_{+}[\hat{O}(t)] = \mathrm{E}\,[\hat{O}(t)]\,:$$

- We can look for $\hat{O}(\,) = \begin{cases} \hat{O}_{+}(\,) & \text{if} \; 2\, X_{+} \\ \hat{O}\;(\,) & \text{if} \; 2\, X \\ 1\, f\;(\,)\quad g & \text{else:} \end{cases}$

Decomposition + change of measure:

$$\mathrm{R}\;(\hat{O}) = \mathrm{E}_{+}\;\frac{(t)}{(\;t)}\;\hat{O}_{+}(t) + \mathrm{E}\;\frac{(t)}{(\;t)}\;\hat{O}\;(t) + !:$$

We should choose $\hat{O}\;(\,) = 1\;\left\{\frac{(t)}{(\;t)}\right\}:$

Optimal fair classification under unawareness

Theorem (G., Schreuder and Chzhen, 2023)

[illegible]

$$
\hat{\ } (\ ) = \begin{cases} 1 \{ [illegible] \} & \text{[illegible]} \in X_+ \\ 1 \{ [illegible] \} & \text{[illegible]} \in X \\ 1 \{ f(\ ) \ [illegible] \} & \text{[illegible]} \end{cases}
$$

[illegible]

$$
\mathrm{P}_+ ([illegible]) = \mathrm{E} ([illegible]) :
$$

Remarks: In the unawareness framework, DP-fair optimal classification with risk measure R

- tries to guess the sensitive attribute;
- **harms** some individuals from **protected group**;
- does not preserves **rational ordering**

Remarks: In the unawareness framework, DP-fair optimal classification with risk measure R

- tries to guess the sensitive attribute;
- **harms** some individuals from **protected group**;
- does not preserves **rational ordering**
- may not preserves **monotonicity**

Remarks: In the unawareness framework, DP-fair optimal classification with risk measure R

- tries to guess the sensitive attribute;
- **harms** some individuals from **protected group**;
- does not preserves **rational ordering**
- may not preserves **monotonicity**

Monotonicity would imply that -almost surely,

$$
{}^{0} =) \begin{cases} \left(' \; 2 \; X_+ : \dfrac{(')}{\underbrace{(')}_{\hat{O}\;(')\;=\;1}} \quad (\;) \right) & \left(' \; 2 \; X_+ : \dfrac{(')}{\underbrace{(')}_{\hat{O}_0(')\;=\;1}}{}^{0} \quad (\; {}^{0}) \right) \\ \left(' \; 2 \; X \; : \dfrac{(')}{\underbrace{(')}_{\hat{O}\;(')\;=\;1}} \quad (\;) \right) & \left(' \; 2 \; X \; : \dfrac{(')}{\underbrace{(')}_{\hat{O}_0(')\;=\;1}}{}^{0} \quad (\; {}^{0}) \right) : \end{cases}
$$

Relationship between classification and regression

Original question: Is 1 $f\grave{\imath}$ (’) g optimal?

- Without monotonicity, no! (The optimal classifier cannot be of the form 1 $f\grave{\imath}$(’) g.)
- **(Divol and G., 2024 - Informal)** With monotonicity, yes!
- Both behaviours can be observed.

Conclusion

In the awareness framework:

- Optimal fair classifier verifies desirable properties : does no harm to the protected group, preserves rational ordering, preserves monotonicity.
- For general performance measures, $\hat{O}$ (’ ; ’) = 1 fì (’ ; ’) g

In the unawareness framework:

- Both the optimal fair classifier and the optimal fair regression function rely on guessing the sensitive attribute;
- Neither of them preserve rational ordering, fair classification harms indivual from the protected group.
- If monotonicity is verified, $\hat{O}$ (’) = 1 fì (’) g
- If monotonicity is not verified, $\hat{O}$ (’) $\neq$ 1 fì (’) g.