Revisiting Fair-PAC Learning and the Axioms of Cardinal Welfare

Cyrus Cousins

University of Massachusetts Amherst
Department of Computer Science

Spring 2023

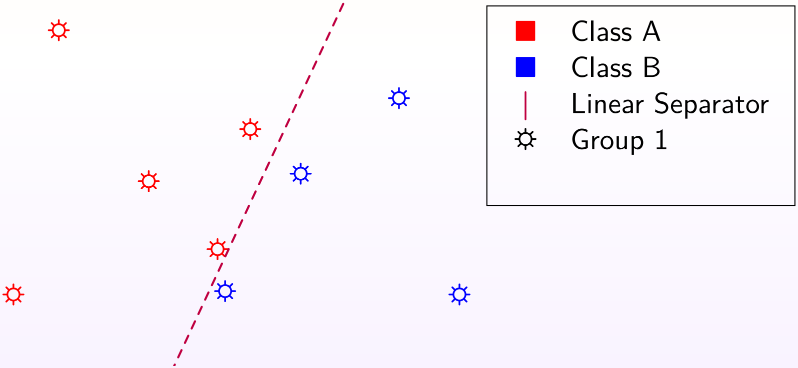
https://www.cyruscousins.online/

Learning Fair Linear Classifiers

- Fair machine learning considers *multiple groups* $(\boldsymbol{x}_{1:g,1:m}, \boldsymbol{y}_{1:g,1:m})$

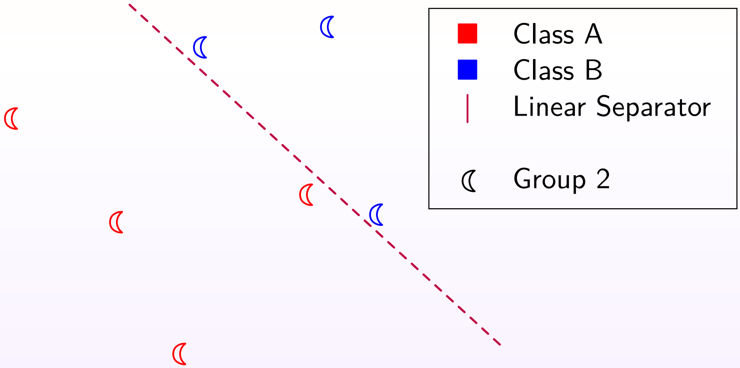
Learning Fair Linear Classifiers

- Fair machine learning considers *multiple groups* $(\boldsymbol{x}_{1:g,1:m}, \boldsymbol{y}_{1:g,1:m})$
- We can handle each group individually
  - *Empirical utility maximization*



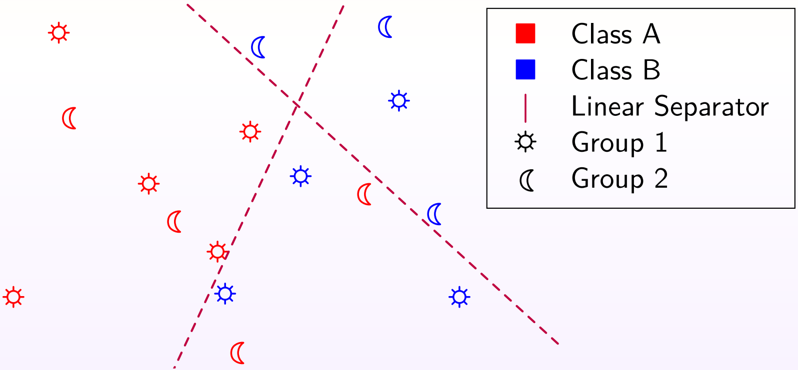
Learning Fair Linear Classifiers

- Fair machine learning considers *multiple groups* $(\boldsymbol{x}_{1:g,1:m}, \boldsymbol{y}_{1:g,1:m})$
- We can handle each group individually
  - *Empirical utility maximization*



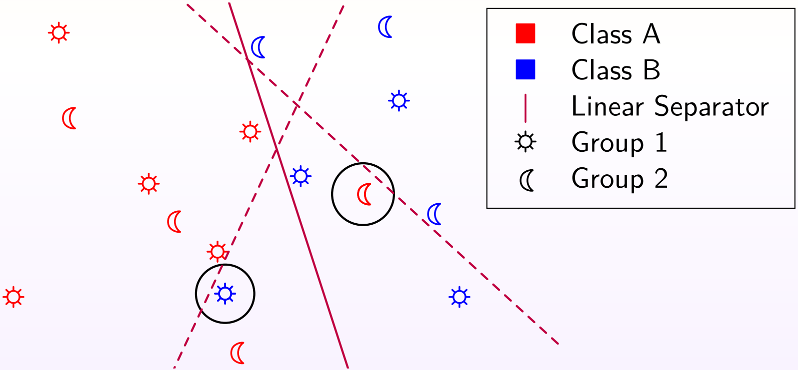
Learning Fair Linear Classifiers

- Fair machine learning considers *multiple groups* $(\boldsymbol{x}_{1:g,1:m}, \boldsymbol{y}_{1:g,1:m})$
- We can handle each group individually
  - *Empirical utility maximization*
- What is the best classifier *overall?*



Learning Fair Linear Classifiers

- Fair machine learning considers *multiple groups* $(\boldsymbol{x}_{1:g,1:m}, \boldsymbol{y}_{1:g,1:m})$
- We can handle each group individually
  - *Empirical utility maximization*
- What is the best classifier *overall?*
  - *Empirical welfare maximization*



Welfare in Machine Learning

- Cardinal welfare functions *mathematically quantify* overall wellbeing
  - *Welfare* $\mathrm{W}(\cdot)$ encodes an *ideal notion* of *societal wellbeing* (fairness)
  - Many reasonable choices available

Welfare in Machine Learning

- Cardinal welfare functions *mathematically quantify* overall wellbeing
  - *Welfare* $\mathrm{W}(\cdot)$ encodes an *ideal notion* of *societal wellbeing* (fairness)
  - Many reasonable choices available
- The *social planning problem*
  - Select allocation of goods and services to *maximize welfare*
  - Fair ML learns a policy from data to *maximize welfare* of *per-group* utilities

Welfare in Machine Learning

- Cardinal welfare functions *mathematically quantify* overall wellbeing
  - *Welfare* $\mathrm{W}(\cdot)$ encodes an *ideal notion* of *societal wellbeing* (fairness)
  - Many reasonable choices available
- The *social planning problem*
  - Select allocation of goods and services to *maximize welfare*
  - Fair ML learns a policy from data to *maximize welfare* of *per-group* utilities
- Welfare maximization in machine learning:
  - Assume a *utility function* $\mathrm{U}(\cdot)$, per-group distributions $\mathcal{D}_1, \dots, \mathcal{D}_g$

$$h^* \doteq \operatorname*{argmax}_{h \in \mathcal{H}} \mathrm{W}\left(\underbrace{\mathop{\mathbb{E}}_{(x,y) \sim \mathcal{D}_1} \left[\mathrm{U}(h(x), y) \right]}_{\text{Group 1 Expected Utility}}, \dots, \underbrace{\mathop{\mathbb{E}}_{(x,y) \sim \mathcal{D}_g} \left[\mathrm{U}(h(x), y) \right]}_{\text{Group } g \text{ Expected Utility}} \right)$$

Welfare in Machine Learning

- Cardinal welfare functions *mathematically quantify* overall wellbeing
  - *Welfare* $\mathrm{W}(\cdot)$ encodes an *ideal notion* of *societal wellbeing* (fairness)
  - Many reasonable choices available
- The *social planning problem*
  - Select allocation of goods and services to *maximize welfare*
  - Fair ML learns a policy from data to *maximize welfare* of *per-group* utilities
- Welfare maximization in machine learning:
  - Assume a *utility function* $\mathrm{U}(\cdot)$, per-group distributions $\mathcal{D}_1, \ldots, \mathcal{D}_g$

$$h^* \doteq \operatorname*{argmax}_{h \in \mathcal{H}} \mathrm{W}\left(\underbrace{\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_1}\left[\mathrm{U}(h(x), y)\right]}_{\text{Group 1 Expected Utility}}, \ldots, \underbrace{\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_g}\left[\mathrm{U}(h(x), y)\right]}_{\text{Group } g \text{ Expected Utility}} \right)$$

⚠ Don't know $\mathcal{D}_{1:g}$; have to work from *training samples* $\boldsymbol{x}_{1:g}, \boldsymbol{y}_{1:g}$

- Select $\hat{h}$ to optimize *empirical welfare*

$$\hat{h} \doteq \operatorname*{argmax}_{h \in \mathcal{H}} \mathrm{W}\left(\mathop{\widehat{\mathbb{E}}}_{(x,y)\in(\boldsymbol{x}_1,\boldsymbol{y}_1)}\left[\mathrm{U}(h(x), y)\right], \ldots, \mathop{\widehat{\mathbb{E}}}_{(x,y)\in(\boldsymbol{x}_g,\boldsymbol{y}_g)}\left[\mathrm{U}(h(x), y)\right] \right)$$

The Power-Mean Welfare Function

Suppose: Positive utility vector $\boldsymbol{u} = \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_g \rangle$ representing *utility* of each group
Positive weights vector $\boldsymbol{w}$ that sums to 1

The Power-Mean Welfare Function

Suppose: Positive utility vector $\boldsymbol{u} = \langle \boldsymbol{u}_1, \dots, \boldsymbol{u}_g \rangle$ representing *utility* of each group
Positive weights vector $\boldsymbol{w}$ that sums to 1

$$\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) \doteq \begin{cases} p \in \mathbb{R}_{\pm} & \sqrt[p]{\sum_{i=1}^{g} \boldsymbol{w}_i \boldsymbol{u}_i^p} \\ & \\ & \end{cases}$$

The Power-Mean Welfare Function

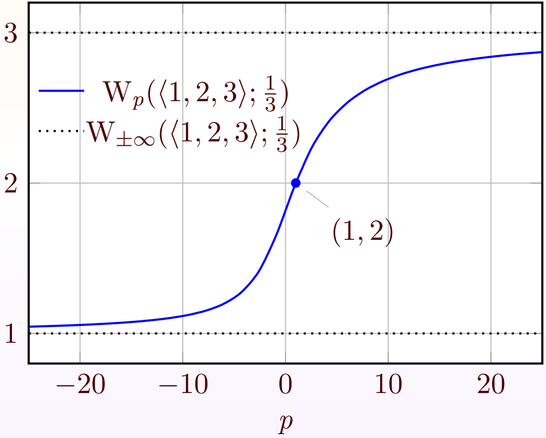
Suppose: Positive utility vector $\boldsymbol{u} = \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_g \rangle$ representing *utility* of each group
Positive weights vector $\boldsymbol{w}$ that sums to 1

$$\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) \doteq \begin{cases} p \in \mathbb{R}_{\pm} & \sqrt[p]{\sum_{i=1}^{g} \boldsymbol{w}_i \boldsymbol{u}_i^p} \\ p = -\infty & \min_{i \in 1, \ldots, g} \boldsymbol{u}_i \\ p = 0 & \prod_{i=1}^{g} \boldsymbol{u}_i^{\boldsymbol{w}_i} \\ p = \infty & \max_{i \in 1, \ldots, g} \boldsymbol{u}_i \end{cases}$$

The Power-Mean Welfare Function

Suppose: Positive utility vector $\boldsymbol{u} = \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_g \rangle$ representing *utility* of each group
Positive weights vector $\boldsymbol{w}$ that sums to 1

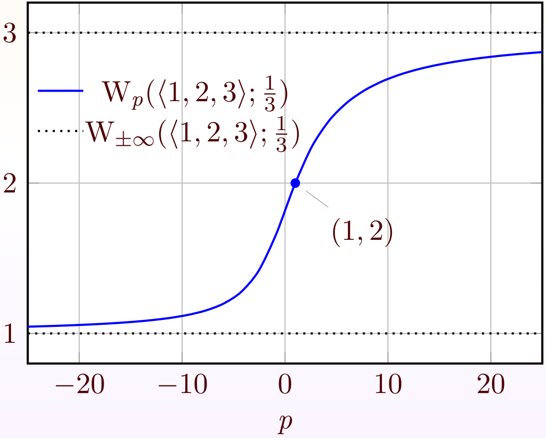
$$\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) \doteq \begin{cases} p \in \mathbb{R}_{\pm} & \sqrt[p]{\sum_{i=1}^{g} \boldsymbol{w}_i \boldsymbol{u}_i^p} \\ p = -\infty & \min_{i \in 1,\ldots,g} \boldsymbol{u}_i \\ p = 0 & \prod_{i=1}^{g} \boldsymbol{u}_i^{\boldsymbol{w}_i} \\ p = \infty & \max_{i \in 1,\ldots,g} \boldsymbol{u}_i \end{cases}$$



The Power-Mean Welfare Function

Suppose: Positive utility vector $\boldsymbol{u} = \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_g \rangle$ representing *utility* of each group
Positive weights vector $\boldsymbol{w}$ that sums to 1

$$\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) \doteq \begin{cases} p \in \mathbb{R}_{\pm} & \sqrt[p]{\sum_{i=1}^{g} \boldsymbol{w}_i \boldsymbol{u}_i^p} \\ p = -\infty & \min_{i \in 1, \ldots, g} \boldsymbol{u}_i \\ p = 0 & \prod_{i=1}^{g} \boldsymbol{u}_i^{\boldsymbol{w}_i} \\ p = \infty & \max_{i \in 1, \ldots, g} \boldsymbol{u}_i \end{cases}$$



- *Smooth interpolation* between *minimum*, *arithmetic mean*, and *maximum*
  - Other special cases: *geometric*, *harmonic*, and *quadratic* means

Continuity of Welfare Functions

- Continuity, coarsely speaking:
  - *Small changes to utility values* $\mapsto$ *small changes to welfare*

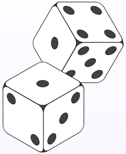
$$\mathrm{W}_p(\boldsymbol{u};\boldsymbol{w}) \approx \mathrm{W}_p(\boldsymbol{u} \pm \varepsilon;\boldsymbol{w})$$

Continuity of Welfare Functions

- Continuity, coarsely speaking:
  - *Small changes to utility values* $\mapsto$ *small changes to welfare*

$$\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) \approx \mathrm{W}_p(\boldsymbol{u} \pm \varepsilon; \boldsymbol{w})$$

- Implications to *philosophy*, *stability*, and *estimation*
  - Notions of *stability* lead to *statistical*, *privacy*, and *robustness* guarantees

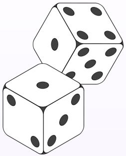


Continuity of Welfare Functions

- Continuity, coarsely speaking:
  - *Small changes to utility values* $\mapsto$ *small changes to welfare*

$$\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) \approx \mathrm{W}_p(\boldsymbol{u} \pm \varepsilon; \boldsymbol{w})$$

- Implications to *philosophy*, *stability*, and *estimation*
  - Notions of *stability* lead to *statistical*, *privacy*, and *robustness* guarantees



- Standard axiom: assume the ε-δ limit definition of continuity for welfare
  - Stronger continuity properties imply stronger guarantees!

A Deep Dive Into Continuity

Definition (Hölder Continuity)

$\mathrm{W}(\boldsymbol{u}; \boldsymbol{w})$ is *Hölder continuous* in $\boldsymbol{u}$ with respect to norm $\|\cdot\|_{\mathrm{W}}$ if there exist

1. scale $\lambda \geq 0$
2. power $\alpha \in (0, 1]$

such that for all $\boldsymbol{u}, \boldsymbol{u}'$, it holds that

$$\left|\mathrm{W}(\boldsymbol{u}; \boldsymbol{w}) - \mathrm{W}(\boldsymbol{u}'; \boldsymbol{w})\right| \leq \lambda \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_{\mathrm{W}}^{\alpha} \enspace .$$

A Deep Dive Into Continuity

Definition (Hölder Continuity)

$\mathrm{W}(\boldsymbol{u}; \boldsymbol{w})$ is *Hölder continuous* in $\boldsymbol{u}$ with respect to norm $\|\cdot\|_{\mathrm{W}}$ if there exist

1. scale $\lambda \geq 0$
2. power $\alpha \in (0, 1]$

such that for all $\boldsymbol{u}, \boldsymbol{u}'$, it holds that

$$\left|\mathrm{W}(\boldsymbol{u}; \boldsymbol{w}) - \mathrm{W}(\boldsymbol{u}'; \boldsymbol{w})\right| \leq \lambda \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_{\mathrm{W}}^{\alpha} \ .$$

Such a function is λ-α-$\|\cdot\|_{\mathrm{W}}$ Hölder continuous.

A Deep Dive Into Continuity

Definition (Hölder Continuity)

$\mathrm{W}(\boldsymbol{u}; \boldsymbol{w})$ is *Hölder continuous* in $\boldsymbol{u}$ with respect to norm $\|\cdot\|_{\mathrm{W}}$ if there exist

1. scale $\lambda \geq 0$
2. power $\alpha \in (0, 1]$

such that for all $\boldsymbol{u}, \boldsymbol{u}'$, it holds that

$$\left|\mathrm{W}(\boldsymbol{u}; \boldsymbol{w}) - \mathrm{W}(\boldsymbol{u}'; \boldsymbol{w})\right| \leq \lambda \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_{\mathrm{W}}^{\alpha} \enspace .$$

Such a function is λ-α-$\|\cdot\|_{\mathrm{W}}$ Hölder continuous.
If $\alpha = 1$, it is λ-$\|\cdot\|_{\mathrm{W}}$ Lipschitz continuous.

A Deep Dive Into Continuity

Definition (Hölder Continuity)

$\mathrm{W}(\boldsymbol{u}; \boldsymbol{w})$ is *Hölder continuous* in $\boldsymbol{u}$ with respect to norm $\|\cdot\|_{\mathrm{W}}$ if there exist

1. scale $\lambda \geq 0$
2. power $\alpha \in (0, 1]$

such that for all $\boldsymbol{u}, \boldsymbol{u}'$, it holds that

$$\left|\mathrm{W}(\boldsymbol{u}; \boldsymbol{w}) - \mathrm{W}(\boldsymbol{u}'; \boldsymbol{w})\right| \leq \lambda \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_{\mathrm{W}}^{\alpha} \enspace .$$

Such a function is λ-α-$\|\cdot\|_{\mathrm{W}}$ Hölder continuous.
If $\alpha = 1$, it is λ-$\|\cdot\|_{\mathrm{W}}$ Lipschitz continuous.

- Lipschitz: bound the impact of *infinitessimal changes*
- Hölder: bound the impact of *small changes*

A Deep Dive Into Continuity

Definition (Hölder Continuity)

$\mathrm{W}(\boldsymbol{u};\boldsymbol{w})$ is *Hölder continuous* in $\boldsymbol{u}$ with respect to norm $\|\cdot\|_{\mathrm{W}}$ if there exist

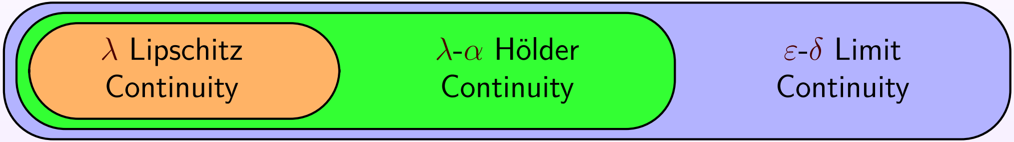
1. scale $\lambda \geq 0$
2. power $\alpha \in (0, 1]$

such that for all $\boldsymbol{u}, \boldsymbol{u}'$, it holds that

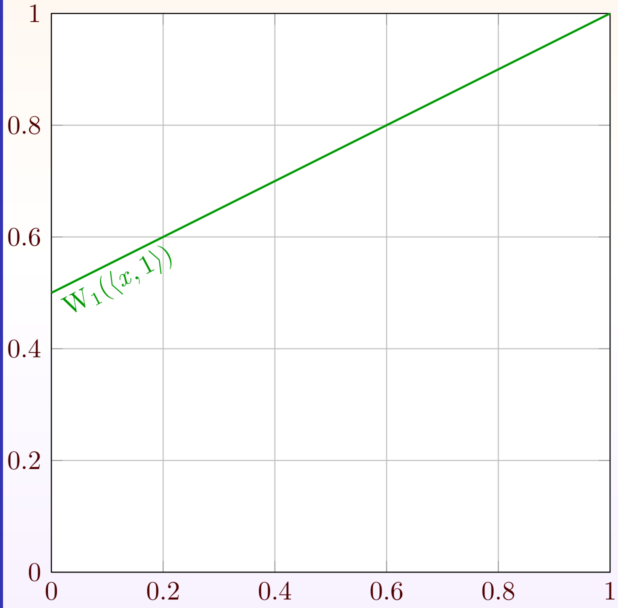
$$\left|\mathrm{W}(\boldsymbol{u};\boldsymbol{w}) - \mathrm{W}(\boldsymbol{u}';\boldsymbol{w})\right| \leq \lambda \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_{\mathrm{W}}^{\alpha} \enspace .$$

Such a function is λ-α-$\|\cdot\|_{\mathrm{W}}$ Hölder continuous.
If $\alpha = 1$, it is λ-$\|\cdot\|_{\mathrm{W}}$ Lipschitz continuous.

- Lipschitz: bound the impact of *infinitessimal changes*
- Hölder: bound the impact of *small changes*



Small-Scale Behavior of Power-Means

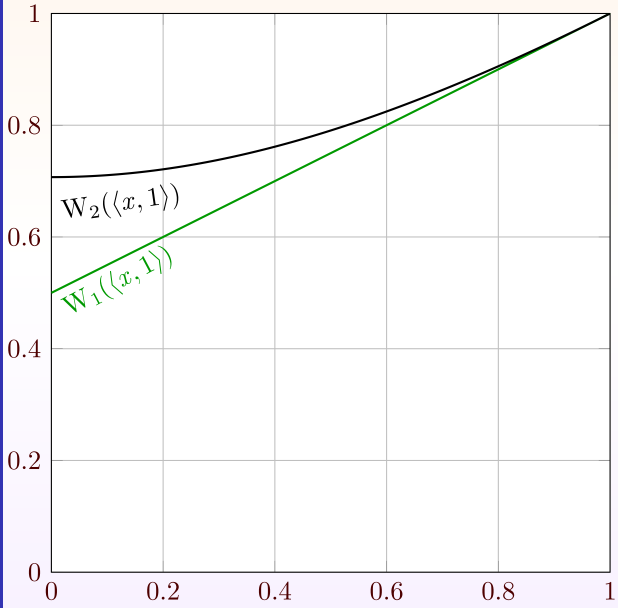


$$\frac{\partial}{\partial \boldsymbol{u}_i} \mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) = \boldsymbol{w}_i \underbrace{\left(\frac{\boldsymbol{u}_i}{\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w})} \right)}_{\text{Relative utility}}^{p-1}$$

Assuming *unit range*

1. $p \geq 1$: 1-$\|\cdot\|_\infty$ Lipschitz

Small-Scale Behavior of Power-Means

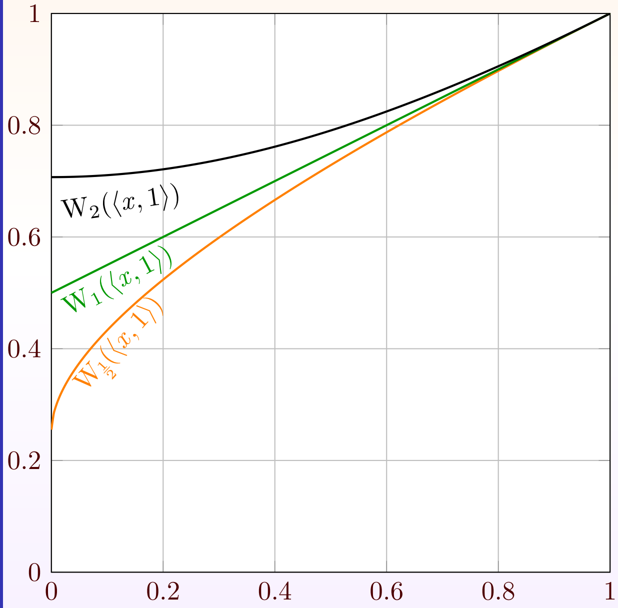


$$\frac{\partial}{\partial \boldsymbol{u}_i} \mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) = \boldsymbol{w}_i \underbrace{\left(\frac{\boldsymbol{u}_i}{\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w})} \right)}_{\text{Relative utility}}^{p-1}$$

Assuming *unit range*

1. $p \geq 1$: 1-$\|\cdot\|_\infty$ Lipschitz

Small-Scale Behavior of Power-Means

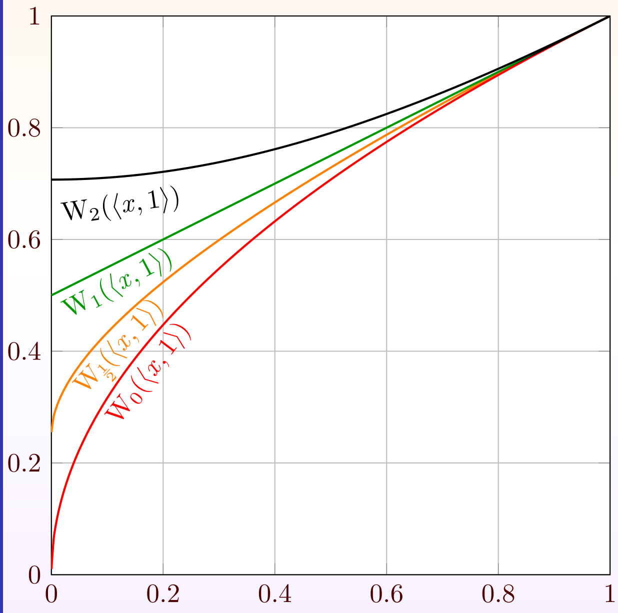


$$\frac{\partial}{\partial \boldsymbol{u}_i}\mathrm{W}_p(\boldsymbol{u};\boldsymbol{w}) = \boldsymbol{w}_i \underbrace{\left(\frac{\boldsymbol{u}_i}{\mathrm{W}_p(\boldsymbol{u};\boldsymbol{w})}\right)}_{\text{Relative utility}}{}^{p-1}$$

Assuming *unit range*

1. $p \geq 1$: 1-$\|\cdot\|_\infty$ Lipschitz
2. $p \in (0, 1)$: $\frac{1}{p}$-p-$\|\cdot\|_\infty$ Hölder

Small-Scale Behavior of Power-Means

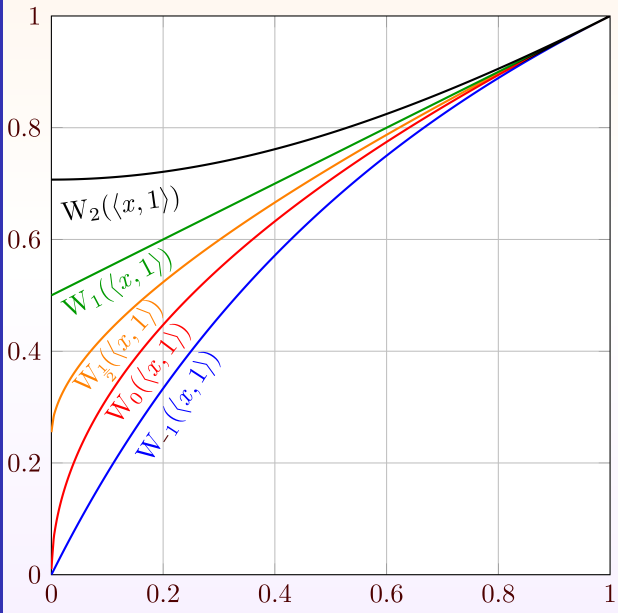


$$\frac{\partial}{\partial \boldsymbol{u}_i}\mathrm{W}_p(\boldsymbol{u};\boldsymbol{w}) = \boldsymbol{w}_i \underbrace{\left(\frac{\boldsymbol{u}_i}{\mathrm{W}_p(\boldsymbol{u};\boldsymbol{w})}\right)}_{\text{Relative utility}}{}^{p-1}$$

Assuming *unit range*

1. $p \geq 1$: 1-$\|\cdot\|_\infty$ Lipschitz
2. $p \in (0,1)$: $\frac{1}{p}$-p-$\|\cdot\|_\infty$ Hölder
3. $p = 0$: 1-$\boldsymbol{w}_{\min}$-$\|\cdot\|_\infty$ Hölder

Small-Scale Behavior of Power-Means

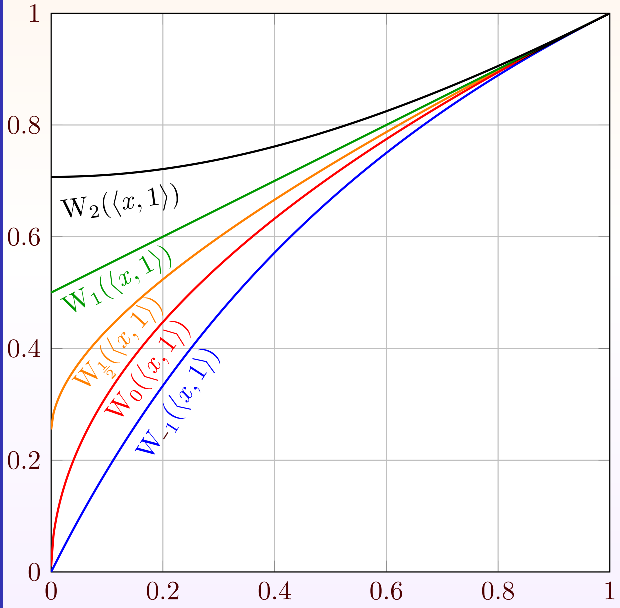


$$\frac{\partial}{\partial \boldsymbol{u}_i} \mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) = \boldsymbol{w}_i \underbrace{\left(\frac{\boldsymbol{u}_i}{\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w})} \right)}_{\text{Relative utility}}{}^{p-1}$$

Assuming *unit range*

1. $p \geq 1$: 1-$\|\cdot\|_\infty$ Lipschitz
2. $p \in (0, 1)$: $\frac{1}{p}$-p-$\|\cdot\|_\infty$ Hölder
3. $p = 0$: 1-$\boldsymbol{w}_{\min}$-$\|\cdot\|_\infty$ Hölder
4. $p < 0$: $\frac{1}{\sqrt[|p|]{\boldsymbol{w}_{\min}}}$-$\|\cdot\|_\infty$ Lipschitz

Small-Scale Behavior of Power-Means



$$\frac{\partial}{\partial \boldsymbol{u}_i} \mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) = \boldsymbol{w}_i \underbrace{\left(\frac{\boldsymbol{u}_i}{\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w})} \right)}_{\text{Relative utility}}{}^{p-1}$$

Assuming *unit range*

1. $p \geq 1$: 1-$\|\cdot\|_\infty$ Lipschitz
2. $p \in (0, 1)$: $\frac{1}{p}$-p-$\|\cdot\|_\infty$ Hölder
3. $p = 0$: 1-$\boldsymbol{w}_{\min}$-$\|\cdot\|_\infty$ Hölder
4. $p < 0$: $\frac{1}{\sqrt[|p|]{\boldsymbol{w}_{\min}}}$-$\|\cdot\|_\infty$ Lipschitz

The “difficult cases:”

- $p \to 0$
- $\boldsymbol{w}_{\min} \to 0$ for $p < 1$

Fair-PAC Learning: CliffsNotes

- PAC Learning
  - Probably approximately optimize *expected utility* over class $\mathcal{H}$
  - Uniformly bound sample complexity of learning
  - Worst-case over distribution $\mathcal{D}$

Fair-PAC Learning: CliffsNotes

- PAC Learning
  - Probably approximately optimize *expected utility* over class $\mathcal{H}$
  - Uniformly bound sample complexity of learning
  - Worst-case over distribution $\mathcal{D}$
- Fair-PAC Learning
  - Probably approximately optimize *welfare* over class $\mathcal{H}$
  - Uniformly bound sample complexity of learning
  - Worst-case over distributions $\mathcal{D}_{1:g}$, welfare functions $\mathrm{W} \in \mathcal{W}$

Fair-PAC Learning

Definition (Fair-PAC Learning)

Suppose

1. hypothesis class $\mathcal{H} \subseteq \mathcal{X} \to \mathcal{Y}'$
2. utility function $\mathrm{U} : \mathcal{Y}' \times \mathcal{Y} \to \mathbb{R}_{0+}$
3. welfare class $\mathcal{W} \subseteq \mathbb{R}_{0+}^{g} \to \mathbb{R}_{0+}$

Fair-PAC Learning

Definition (Fair-PAC Learning)

Suppose

1. hypothesis class $\mathcal{H} \subseteq \mathcal{X} \to \mathcal{Y}'$
2. utility function $\mathrm{U} : \mathcal{Y}' \times \mathcal{Y} \to \mathbb{R}_{0+}$
3. welfare class $\mathcal{W} \subseteq \mathbb{R}_{0+}^{g} \to \mathbb{R}_{0+}$

$\mathcal{H}$ is *fair-PAC-learnable* if there exists an algorithm $\mathcal{A}$ such that for any

1. distributions $\mathcal{D}_{1:g}$ over $(\mathcal{X} \times \mathcal{Y})$
2. welfare function $\mathrm{W}(\cdot; \boldsymbol{w}) \in \mathcal{W}$
3. additive error $\varepsilon > 0$
4. failure probability $\delta \in (0, 1)$

Fair-PAC Learning

Definition (Fair-PAC Learning)

Suppose

1. hypothesis class $\mathcal{H} \subseteq \mathcal{X} \to \mathcal{Y}'$
2. utility function $\mathrm{U} : \mathcal{Y}' \times \mathcal{Y} \to \mathbb{R}_{0+}$
3. welfare class $\mathcal{W} \subseteq \mathbb{R}^{g}_{0+} \to \mathbb{R}_{0+}$

$\mathcal{H}$ is *fair-PAC-learnable* if there exists an algorithm $\mathcal{A}$ such that for any

1. distributions $\mathcal{D}_{1:g}$ over $(\mathcal{X} \times \mathcal{Y})$
2. welfare function $\mathrm{W}(\cdot; \boldsymbol{w}) \in \mathcal{W}$
3. additive error $\varepsilon > 0$
4. failure probability $\delta \in (0, 1)$

$\mathcal{A}$ can identify a hypothesis $\hat{h} \in \mathcal{H}$ such that

1. $\mathcal{A}$ has $\mathrm{m}_{\mathcal{W},\mathcal{H}}(\varepsilon, \delta, \mathrm{W}, g)$ sample complexity (per-group)
2. with probability at least $1 - \delta$, $\hat{h}$ obeys

$$\underbrace{\mathrm{W}\left(\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_1} \left[\mathrm{U}(\hat{h}(x), y) \right], \ldots; \boldsymbol{w} \right)}_{\text{Learned model welfare}} \geq \underbrace{\operatorname*{argmax}_{h^* \in \mathcal{H}} \mathrm{W}\left(\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_1} \left[\mathrm{U}(h^*(x), y) \right], \ldots; \boldsymbol{w} \right)}_{\text{Optimal model welfare}} - \varepsilon$$

Uniform Convergence

- Suppose that for a sample $(\boldsymbol{x}, \boldsymbol{y})$ of size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$, it holds that

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta \enspace .$$

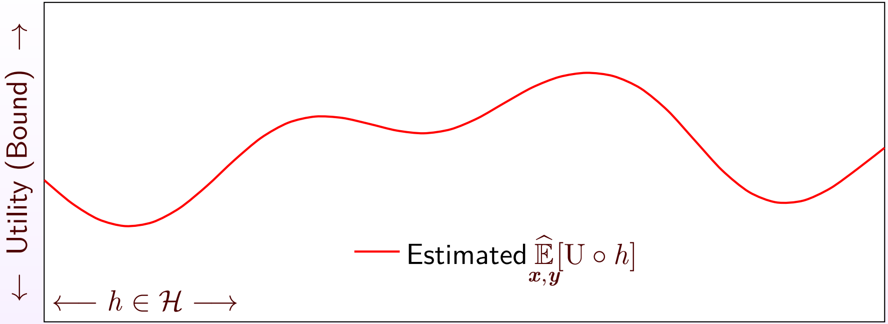
- Many ways to show this:
  - Vapnik-Chervonenkis dimension
  - Rademacher averages

Uniform Convergence

- Suppose that for a sample $(\boldsymbol{x}, \boldsymbol{y})$ of size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$, it holds that

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}}\left|\mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h]\right| > \varepsilon\right) < \delta \enspace .$$

- Many ways to show this:
  - Vapnik-Chervonenkis dimension
  - Rademacher averages

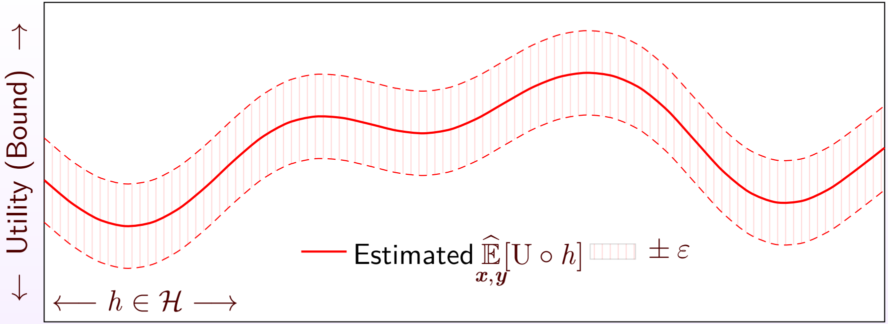


Uniform Convergence

- Suppose that for a sample $(\boldsymbol{x}, \boldsymbol{y})$ of size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$, it holds that

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta \enspace .$$

- Many ways to show this:
  - Vapnik-Chervonenkis dimension
  - Rademacher averages

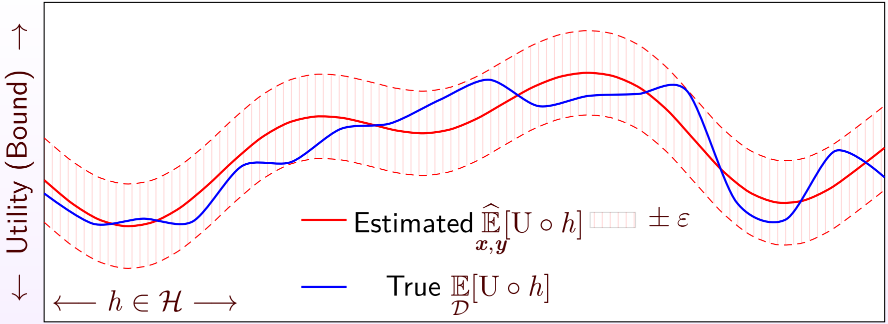


Uniform Convergence

- Suppose that for a sample $(\boldsymbol{x}, \boldsymbol{y})$ of size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$, it holds that

$$\mathbb{P}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathbb{E}_{\mathcal{D}}[\mathrm{U} \circ h] - \widehat{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta \enspace .$$

- Many ways to show this:
  - Vapnik-Chervonenkis dimension
  - Rademacher averages

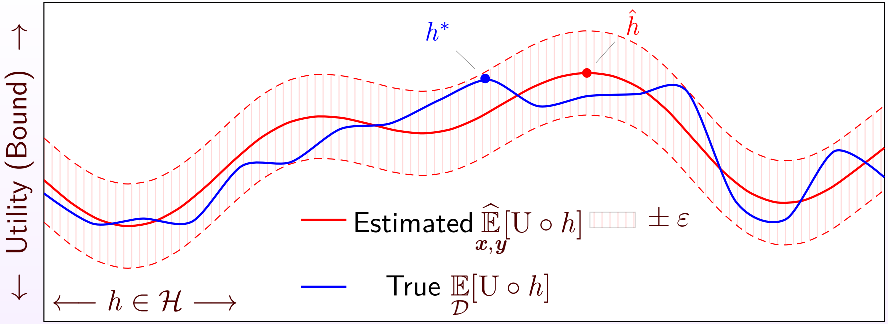


Uniform Convergence

- Suppose that for a sample $(\boldsymbol{x}, \boldsymbol{y})$ of size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$, it holds that

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta \enspace .$$

- Many ways to show this:
  - Vapnik-Chervonenkis dimension
  - Rademacher averages



Uniform Convergence and FPAC Learning

Uniform Convergence:

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h\in\mathcal{H}}\left|\mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U}\circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U}\circ h]\right| > \varepsilon\right) < \delta$$

- What does *uniform convergence* give us?
  - *Asymptotic consistency* of empirical utility maximizer $\hat{h}$
  - UC implies PAC with sample complexity $\mathrm{m}_{\mathcal{H}}(\frac{\varepsilon}{2}, \delta)$

Uniform Convergence and FPAC Learning

Uniform Convergence:

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h\in\mathcal{H}}\left|\mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U}\circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U}\circ h]\right| > \varepsilon\right) < \delta$$

- What does *uniform convergence* give us?
  - *Asymptotic consistency* of empirical utility maximizer $\hat{h}$
  - UC implies PAC with sample complexity $\mathrm{m}_{\mathcal{H}}(\frac{\varepsilon}{2}, \delta)$
- By ε-δ limit continuity of $\mathrm{W}(\cdot; \boldsymbol{w})$ alone:
  - Consistency of empirical welfare maximizer $\hat{h}$
  - ⚠ Convergence rate depends on welfare function

Uniform Convergence and FPAC Learning

Uniform Convergence:

$$\mathbb{P}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h\in\mathcal{H}}\left|\mathbb{E}_{\mathcal{D}}[\mathrm{U}\circ h] - \widehat{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U}\circ h]\right| > \varepsilon\right) < \delta$$

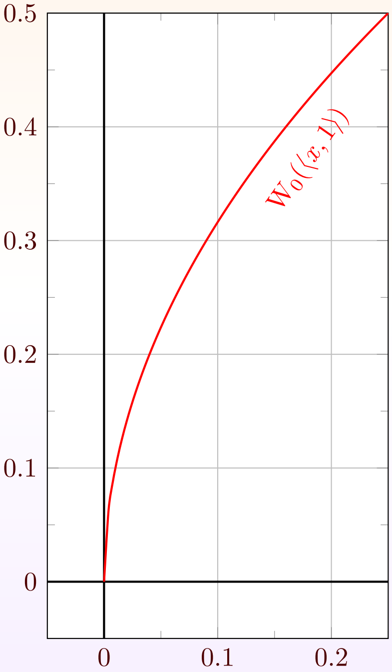
- What does *uniform convergence* give us?
  - *Asymptotic consistency* of empirical utility maximizer $\hat{h}$
  - UC implies PAC with sample complexity $\mathrm{m}_{\mathcal{H}}(\frac{\varepsilon}{2}, \delta)$
- By ε-δ limit continuity of $\mathrm{W}(\cdot; \boldsymbol{w})$ alone:
  - Consistency of empirical welfare maximizer $\hat{h}$
  - ⚠ Convergence rate depends on welfare function
- By how much does $\mathrm{W}(\cdot; \boldsymbol{w})$ magnify error?
  - Hölder continuity analysis

Uniform Convergence and FPAC Learning

Uniform Convergence:

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta$$

- What does *uniform convergence* give us?
  - *Asymptotic consistency* of empirical utility maximizer $\hat{h}$
  - UC implies PAC with sample complexity $\mathrm{m}_{\mathcal{H}}(\frac{\varepsilon}{2}, \delta)$
- By ε-δ limit continuity of $\mathrm{W}(\cdot; \boldsymbol{w})$ alone:
  - Consistency of empirical welfare maximizer $\hat{h}$
  - ⚠ Convergence rate depends on welfare function
- By how much does $\mathrm{W}(\cdot; \boldsymbol{w})$ magnify error?
  - Hölder continuity analysis

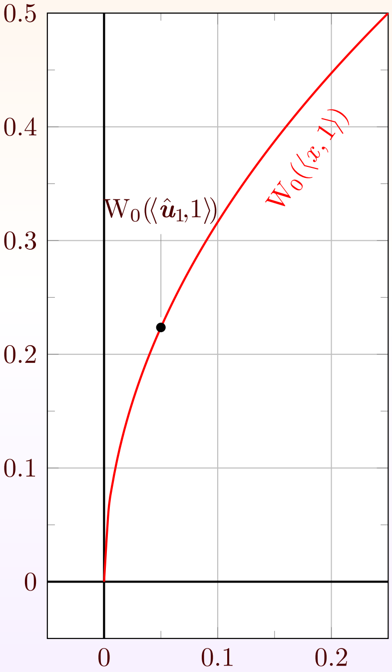


Uniform Convergence and FPAC Learning

Uniform Convergence:

$$\mathbb{P}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h\in\mathcal{H}}\left|\mathbb{E}_{\mathcal{D}}[\mathrm{U}\circ h] - \widehat{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U}\circ h]\right| > \varepsilon\right) < \delta$$

- What does *uniform convergence* give us?
  - *Asymptotic consistency* of empirical utility maximizer $\hat{h}$
  - UC implies PAC with sample complexity $\mathrm{m}_{\mathcal{H}}(\frac{\varepsilon}{2}, \delta)$
- By ε-δ limit continuity of $\mathrm{W}(\cdot; \boldsymbol{w})$ alone:
  - Consistency of empirical welfare maximizer $\hat{h}$
  - ⚠ Convergence rate depends on welfare function
- By how much does $\mathrm{W}(\cdot; \boldsymbol{w})$ magnify error?
  - Hölder continuity analysis

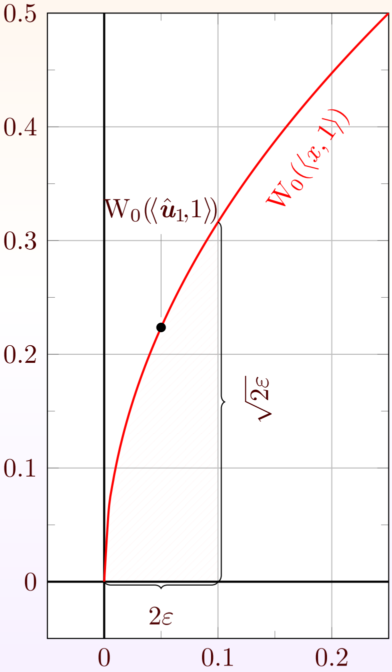


Uniform Convergence and FPAC Learning

Uniform Convergence:

$$\mathbb{P}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h\in\mathcal{H}}\left|\mathbb{E}_{\mathcal{D}}[\mathrm{U}\circ h] - \widehat{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U}\circ h]\right| > \varepsilon\right) < \delta$$

- What does *uniform convergence* give us?
  - *Asymptotic consistency* of empirical utility maximizer $\hat{h}$
  - UC implies PAC with sample complexity $\mathrm{m}_{\mathcal{H}}(\frac{\varepsilon}{2}, \delta)$
- By ε-δ limit continuity of $\mathrm{W}(\cdot;\boldsymbol{w})$ alone:
  - Consistency of empirical welfare maximizer $\hat{h}$
  - ⚠ Convergence rate depends on welfare function
- By how much does $\mathrm{W}(\cdot;\boldsymbol{w})$ magnify error?
  - Hölder continuity analysis



Characterizing Fair-PAC Learnability

Theorem (Welfare Estimation Sample Complexity)

Suppose uniform convergence, i.e., that for sample size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$*, it holds that*

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta \ .$$

Characterizing Fair-PAC Learnability

Theorem (Welfare Estimation Sample Complexity)

Suppose uniform convergence, i.e., that for sample size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$*, it holds that*

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta \enspace .$$

Then $\mathcal{H}$ *is FPAC-learnable with sample complexity*

$$\mathrm{m}_{\mathcal{W},\mathcal{H}}(\varepsilon, \delta, \mathrm{W}, g) \leq \mathrm{m}_{\mathcal{H}}\left(\sqrt[\alpha]{\frac{\varepsilon}{2\lambda}}, \frac{\delta}{g}\right) \enspace .$$

Characterizing Fair-PAC Learnability

Theorem (Welfare Estimation Sample Complexity)

Suppose uniform convergence, i.e., that for sample size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$*, it holds that*

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta \enspace .$$

Then $\mathcal{H}$ *is FPAC-learnable with sample complexity*

$$\mathrm{m}_{\mathcal{W},\mathcal{H}}(\varepsilon, \delta, \mathrm{W}, g) \leq \mathrm{m}_{\mathcal{H}}\left(\sqrt[\alpha]{\frac{\varepsilon}{2\lambda}}, \frac{\delta}{g}\right) \enspace .$$

- Sample complexity of ε-δ learning is usually $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta) \in \mathbf{O}\left(\frac{\ln \frac{1}{\delta}}{\varepsilon^2}\right)$

Characterizing Fair-PAC Learnability

Theorem (Welfare Estimation Sample Complexity)

Suppose uniform convergence, i.e., that for sample size $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta)$, it holds that

$$\mathop{\mathbb{P}}_{\boldsymbol{x},\boldsymbol{y}}\left(\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{\mathcal{D}}[\mathrm{U} \circ h] - \mathop{\widehat{\mathbb{E}}}_{\boldsymbol{x},\boldsymbol{y}}[\mathrm{U} \circ h] \right| > \varepsilon \right) < \delta \enspace .$$

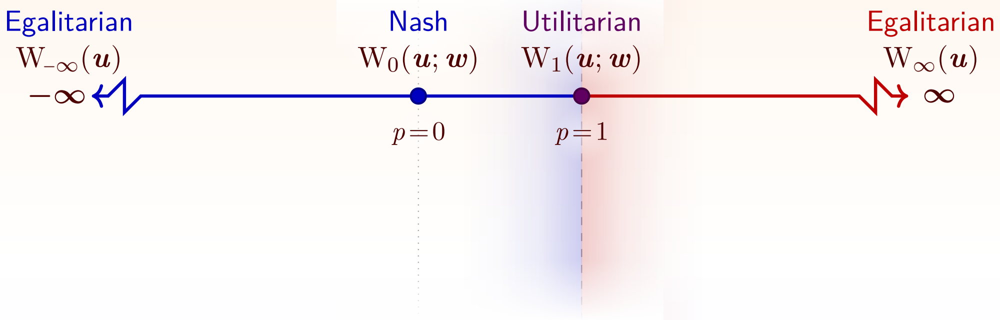
Then $\mathcal{H}$ is FPAC-learnable with sample complexity

$$\mathrm{m}_{\mathcal{W},\mathcal{H}}(\varepsilon, \delta, \mathrm{W}, g) \leq \mathrm{m}_{\mathcal{H}}\left(\sqrt[\alpha]{\frac{\varepsilon}{2\lambda}}, \frac{\delta}{g}\right) \enspace .$$

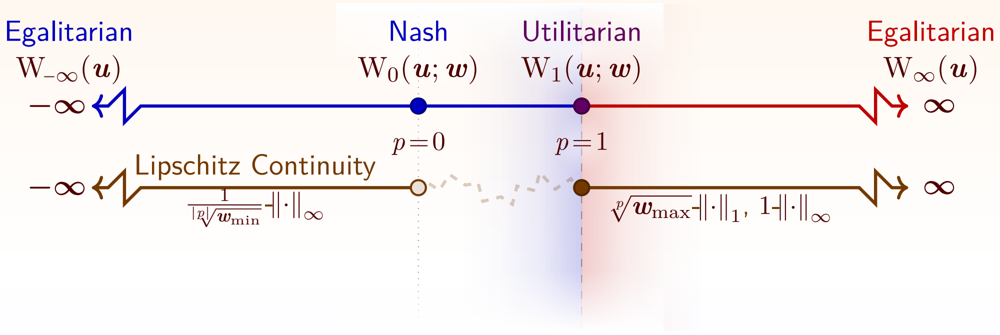
- Sample complexity of ε-δ learning is usually $\mathrm{m}_{\mathcal{H}}(\varepsilon, \delta) \in \mathbf{O}\left(\frac{\ln \frac{1}{\delta}}{\varepsilon^2}\right)$
- Fair-learning the class of *all weighted power-means*:

$$\mathrm{m}_{\mathcal{W},\mathcal{H}}(\varepsilon, \delta, \mathrm{W}, g) \leq \mathrm{m}_{\mathcal{H}}\left(\sqrt[\alpha]{\frac{\varepsilon}{2\lambda}}, \frac{\delta}{g}\right) \in \mathbf{O}\left(\frac{\lambda^{\frac{2}{\alpha}} \ln \frac{g}{\delta}}{\varepsilon^{\frac{2}{\alpha}}}\right) \subseteq \mathbf{O}\left(\frac{\ln \frac{g}{\delta}}{\varepsilon^{\frac{2}{w_{\min}}}}\right)$$

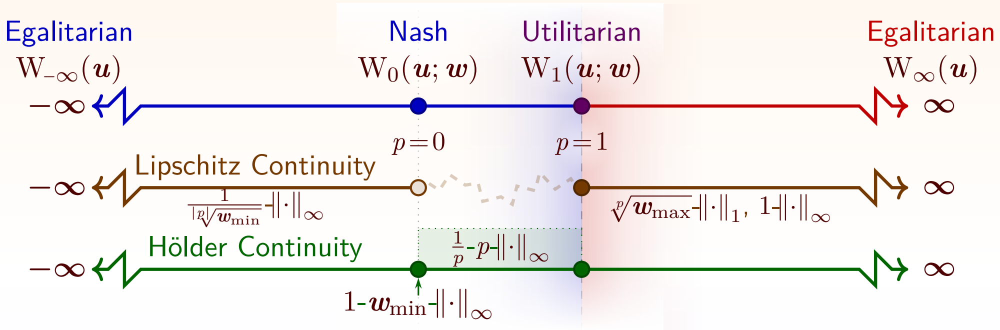
Axiomatic Characterization of Welfare Classes



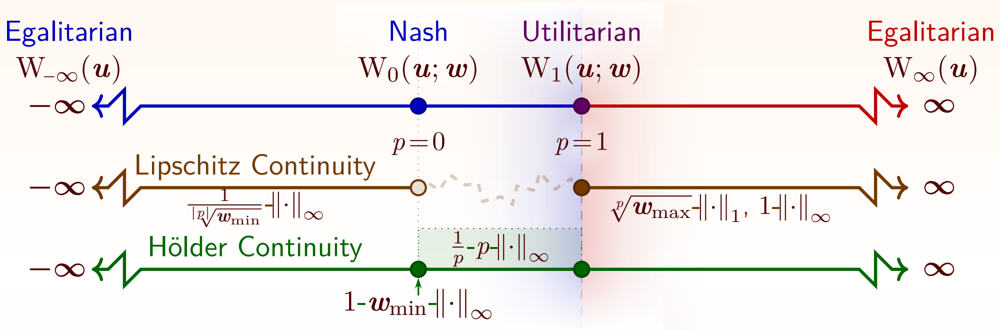
Axiomatic Characterization of Welfare Classes



Axiomatic Characterization of Welfare Classes

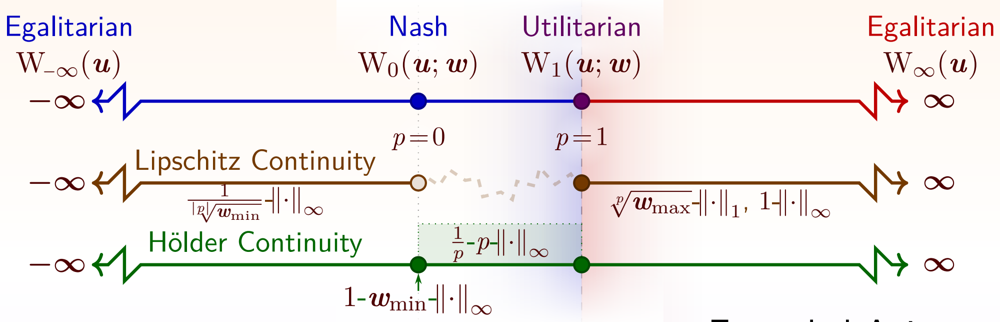


Axiomatic Characterization of Welfare Classes



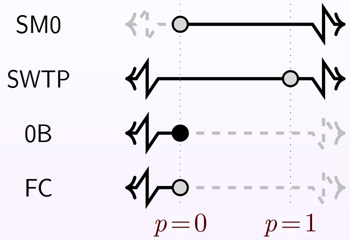
- Axiomatic characterization of welfare functions
  - Uniquely satisfied by $\mathrm{W}_p(\cdot;\boldsymbol{w})$ for $p \leq 1$

Axiomatic Characterization of Welfare Classes

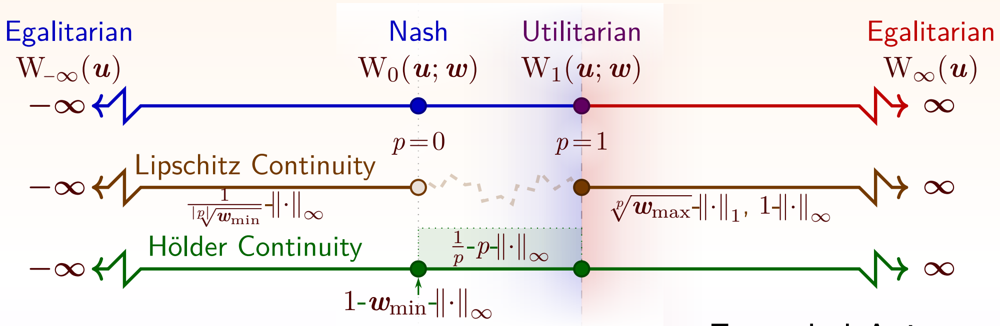


- Axiomatic characterization of welfare functions
  - Uniquely satisfied by $\mathrm{W}_p(\cdot;\boldsymbol{w})$ for $p \leq 1$
- Additional axioms further restrict p
  - FPAC learning efficiency varies by region!

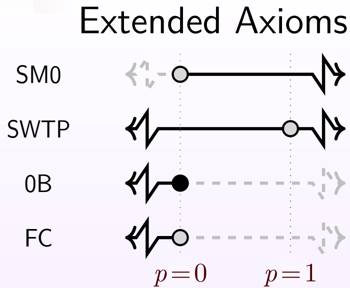
Extended Axioms



Axiomatic Characterization of Welfare Classes



- Axiomatic characterization of welfare functions
  - Uniquely satisfied by $\mathrm{W}_p(\cdot; \boldsymbol{w})$ for $p \leq 1$
- Additional axioms further restrict p
  - FPAC learning efficiency varies by region!
- $p \geq 1$ used similarly for *malfare* and *disutility*
  - Studied in prior work
  - Entire class is uniformly FPAC-learnable (Lipschitz)



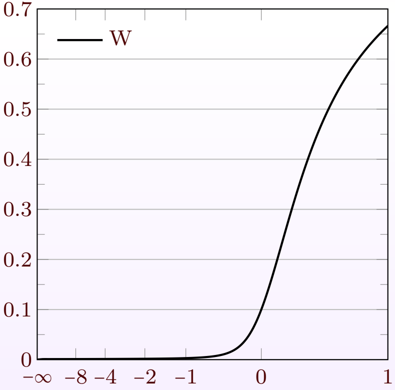
Bandit Experiments

- The Multi-Group One-Armed Bandit
  - Each arm-pull contributes some utility to each of two groups
  - Estimate the *welfare* of the *expected utilities* of each group
- Unless otherwise noted:
  - $\boldsymbol{u} = \langle 0.999, 0.001 \rangle$
  - $\boldsymbol{w} = \langle \frac{2}{3}, \frac{1}{3} \rangle$
  - $m = 100$ samples
  - $p = 0$

Bandit Experiments

- The Multi-Group One-Armed Bandit
  - Each arm-pull contributes some utility to each of two groups
  - Estimate the *welfare* of the *expected utilities* of each group
- Unless otherwise noted:
  - $\boldsymbol{u} = \langle 0.999, 0.001 \rangle$
  - $\boldsymbol{w} = \langle \frac{2}{3}, \frac{1}{3} \rangle$
  - $m = 100$ samples
  - $p = 0$

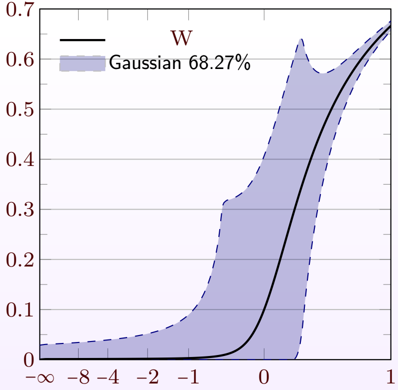
Welfare vs. $p \in [-\infty, 1]$



Bandit Experiments

- The Multi-Group One-Armed Bandit
  - Each arm-pull contributes some utility to each of two groups
  - Estimate the *welfare* of the *expected utilities* of each group
- Unless otherwise noted:
  - $\boldsymbol{u} = \langle 0.999, 0.001 \rangle$
  - $\boldsymbol{w} = \langle \frac{2}{3}, \frac{1}{3} \rangle$
  - $m = 100$ samples
  - $p = 0$

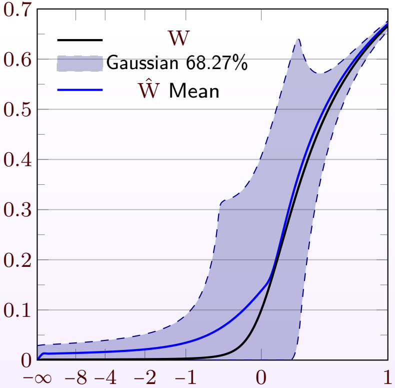
Welfare vs. $p \in [-\infty, 1]$



Bandit Experiments

- The Multi-Group One-Armed Bandit
  - Each arm-pull contributes some utility to each of two groups
  - Estimate the *welfare* of the *expected utilities* of each group
- Unless otherwise noted:
  - $\boldsymbol{u} = \langle 0.999, 0.001 \rangle$
  - $\boldsymbol{w} = \langle \frac{2}{3}, \frac{1}{3} \rangle$
  - $m = 100$ samples
  - $p = 0$

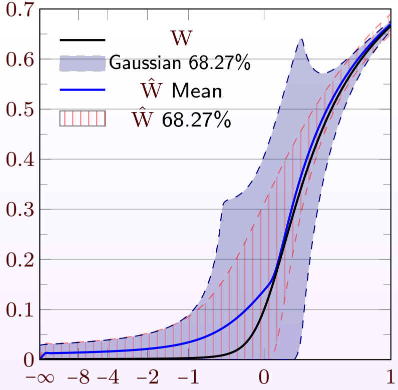
Welfare vs. $p \in [-\infty, 1]$



Bandit Experiments

- The Multi-Group One-Armed Bandit
  - Each arm-pull contributes some utility to each of two groups
  - Estimate the *welfare* of the *expected utilities* of each group
- Unless otherwise noted:
  - $\boldsymbol{u} = \langle 0.999, 0.001 \rangle$
  - $\boldsymbol{w} = \langle \frac{2}{3}, \frac{1}{3} \rangle$
  - $m = 100$ samples
  - $p = 0$

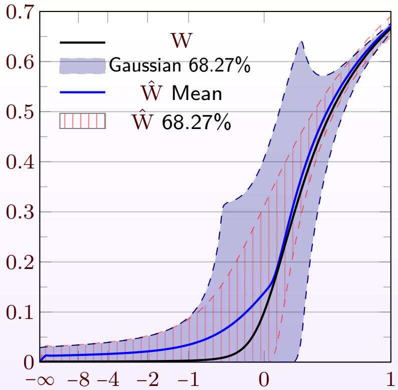
Welfare vs. $p \in [-\infty, 1]$



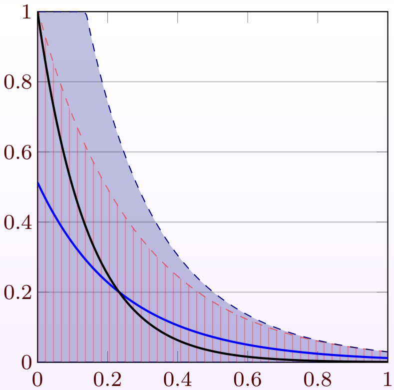
Bandit Experiments

- The Multi-Group One-Armed Bandit
  - Each arm-pull contributes some utility to each of two groups
  - Estimate the *welfare* of the *expected utilities* of each group
- Unless otherwise noted:
  - $\boldsymbol{u} = \langle 0.999, 0.001 \rangle$
  - $\boldsymbol{w} = \langle \frac{2}{3}, \frac{1}{3} \rangle$
  - $m = 100$ samples
  - $p = 0$

Welfare vs. $p \in [-\infty, 1]$



Welfare vs. minority weight $\boldsymbol{w}_2$



Conclusion

- Axiomatically characterize class of fair welfare functions
  - Act as *objective metric* of *subjective utility*
  - Fairness (welfare) varies interpersonally
  - “Reasonable axioms” describe “reasonable people”

$$\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) \doteq \sqrt[p]{\sum_{i=1}^{g} \boldsymbol{w}_i \boldsymbol{u}_i^p}$$

Conclusion

- Axiomatically characterize class of fair welfare functions
  - Act as *objective metric* of *subjective utility*
  - Fairness (welfare) varies interpersonally
  - "Reasonable axioms" describe "reasonable people"

$$\mathrm{W}_p(\boldsymbol{u};\boldsymbol{w}) \doteq \sqrt[p]{\sum_{i=1}^{g} \boldsymbol{w}_i \boldsymbol{u}_i^p}$$

- Analyze continuity properties of fair welfare functions
  - Lipschitz and Hölder continuity

$$\left|\mathrm{W}(\boldsymbol{u};\boldsymbol{w}) - \mathrm{W}(\boldsymbol{u}';\boldsymbol{w})\right| \leq \lambda \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_{\mathrm{W}}^{\alpha}$$

Conclusion

- Axiomatically characterize class of fair welfare functions
  - Act as *objective metric* of *subjective utility*
  - Fairness (welfare) varies interpersonally
  - "Reasonable axioms" describe "reasonable people"

$$\mathrm{W}_p(\boldsymbol{u}; \boldsymbol{w}) \doteq \sqrt[p]{\sum_{i=1}^{g} \boldsymbol{w}_i \boldsymbol{u}_i^p}$$

- Analyze continuity properties of fair welfare functions
  - Lipschitz and Hölder continuity

$$\left|\mathrm{W}(\boldsymbol{u}; \boldsymbol{w}) - \mathrm{W}(\boldsymbol{u}'; \boldsymbol{w})\right| \leq \lambda \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_{\mathrm{W}}^{\alpha}$$

- Fair-PAC learnability for all welfare functions W in class $\mathcal{W}$
  - Uniform convergence implies FPAC-Learnability
  - Polynomial sample complexity preserved except as $p \to 0$ or $\boldsymbol{w}_{\min} \to 0$

$$\mathrm{m}_{\mathcal{W},\mathcal{H}}(\varepsilon, \delta, \mathrm{W}, g) \leq \mathrm{m}_{\mathcal{H}}\left(\sqrt[\alpha]{\frac{\varepsilon}{2\lambda}}, \frac{\delta}{g}\right)$$

Fin