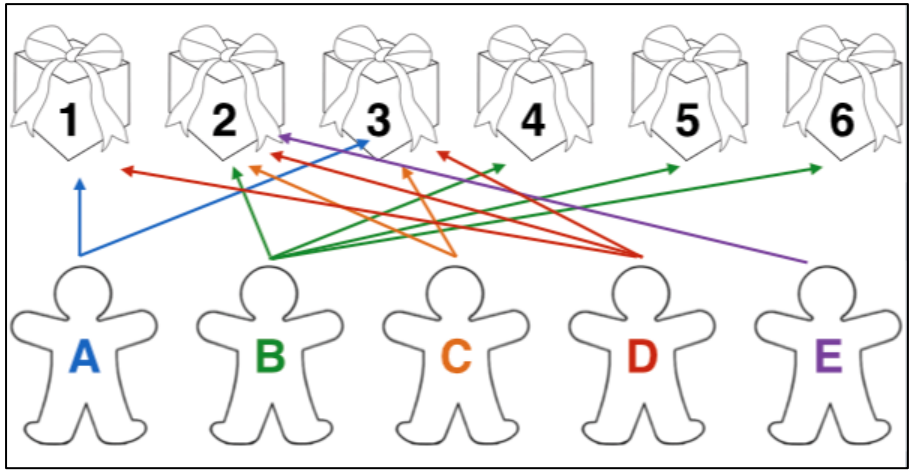
Biased By Design

- The National Resident Matching Program



Hospital residency programs

Graduating medical students

  - Gale-Shapley
  - Implementation choice: program-proposing vs applicant-proposing

Image credit:

Our Foci

- Binary classification algorithms
  - Classify each person as positive (*eg*, high risk for disease) or negative (low risk)
- Scoring functions $f(x) = p \in [0,1]$
  - Often interpreted as a probability; more about that later

Theory of Algorithmic Fairness

- Definitions: Group vs Individual

  Group notions fail under scrutiny

- Group Fairness Examples
  - *Statistical parity*: demographics of accepted students are same as in population
    - 48.7% female
  - *Balance for positive class*: the average score for a positive member of A is the same as the average score for a positive member of B

DworkHardtPitassiReingoldZemel2012

Theory of Algorithmic Fairness

- Definitions: Group vs Individual
  - *Group notions fail under scrutiny*
  - *– steak ads for vegetarians*
  - *– very different distributions, reward minority that "look like" majority*
  - *– surprisingly hard to test*
  - *– natural desiderata are mutually exclusive*
  - *– which groups? Intersectionality?*

- Group Fairness Examples
  - *Statistical parity*: demographics of accepted students are same as in population
  - *Balance for positive class*: the average score for a positive member of A is the same as the average score for a positive member of B

NeilWinship2019

Chouldechova 2016; KleinbergMullainathanRaghavan2016

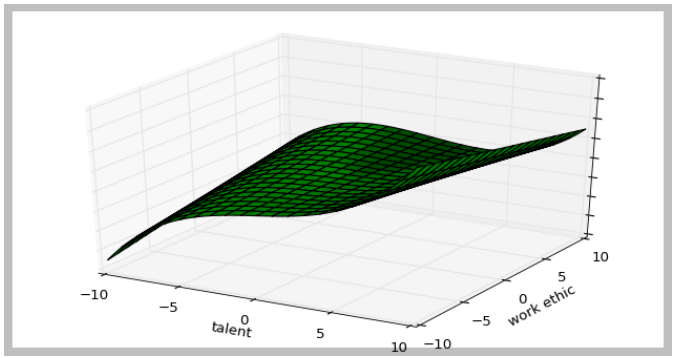
DworkHardtPitassiReingoldZemel2012

Theory of Algorithmic Fairness

- Definitions: Group vs Individual

  Group notions fail under scrutiny

  Individual Fairness requires a task-specific metric



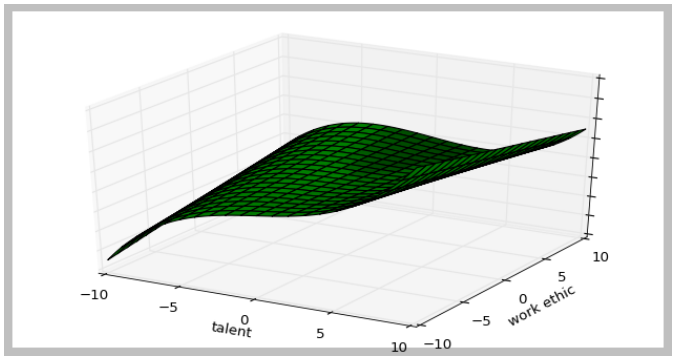
- Individual Fairness
  - People who are similar with respect to a given classification task should be treated similarly
    - $\left||C(x) - C(z)|\right| \leq d_T(x, z)$

Theory of Algorithmic Fairness

- Definitions: Group vs Individual

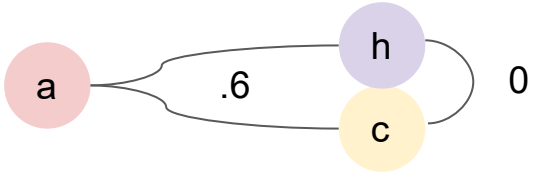
  Group notions fail under scrutiny

  Individual Fairness requires a task-specific metric



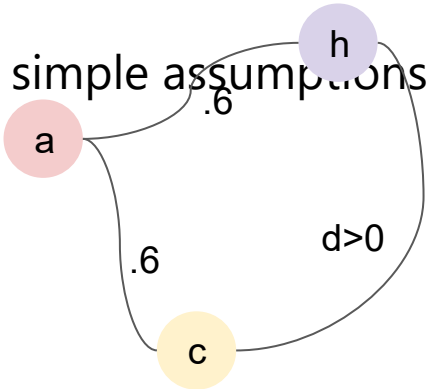
- Individual Fairness
  - People who are similar with respect to a given classification task should be treated similarly
    - $||C(x) - C(z)|| \leq d_T(x, z)$
    - Strong legal foundation
      - $d_T(x, z)$?
    - Ilvento19: O(1) hard queries
    - GillenJungRothKearns18
    - KimReingoldRothblum18
    - RothblumYona18

Tantalizing Breakthrough!



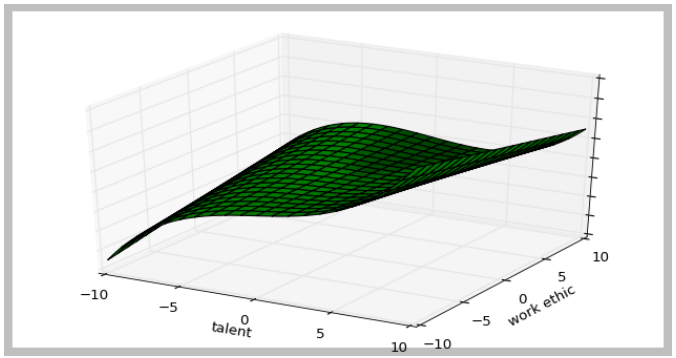
Three Insights for Metric Learning for Individual Fairness:

1. Distances from a single “representative” element produce useful approximations to the true metric.
2. “Parallax” can be achieved by aggregating approximations obtained from a small number of additional representatives
3. Can generalize to unseen elements under simple assumptions about learning threshold functions



Theory of Algorithmic Fairness

- Definitions: Group vs Individual
  - *Group notions fail under scrutiny*
  - *Individual Fairness requires a task-specific metric*



- Individual Fairness
  - People who are similar with respect to a given classification task should be treated similarly
    - Strong legal foundation
      - $d_T(x, z)$?
    - The "Metric Conjecture": a metric can be *extracted* from any "fair" system or "fairness" oracle

DworkIlventoRothblumSur2020

Individual Probabilities: the Defining Problem of AI

Risk predictors assign numbers in [0,1] to individual instances:

- What is the probability that it will rain *tomorrow*?
- What is the probability that *X* will repay the loan?
- What is the probability that *this* tumor will metastasize?

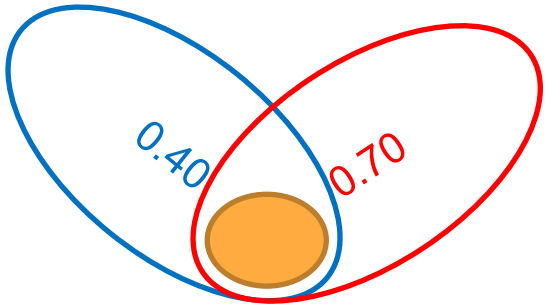
What is the "probability" of a non-repeatable event?

The Tumor Example

- "Probabilities" are learned from binary outcomes data
  – did vs did not metastasize

Locations considered in Study 1

Locations considered in Study 2

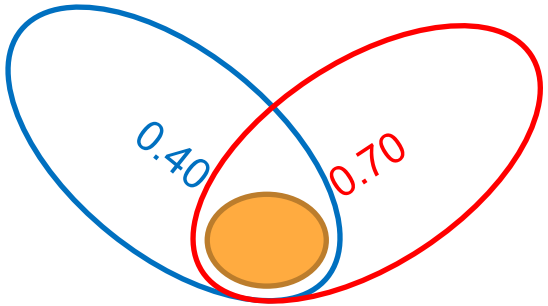


The Tumor Example

- Representation matters!
  – vector for introduction of bias

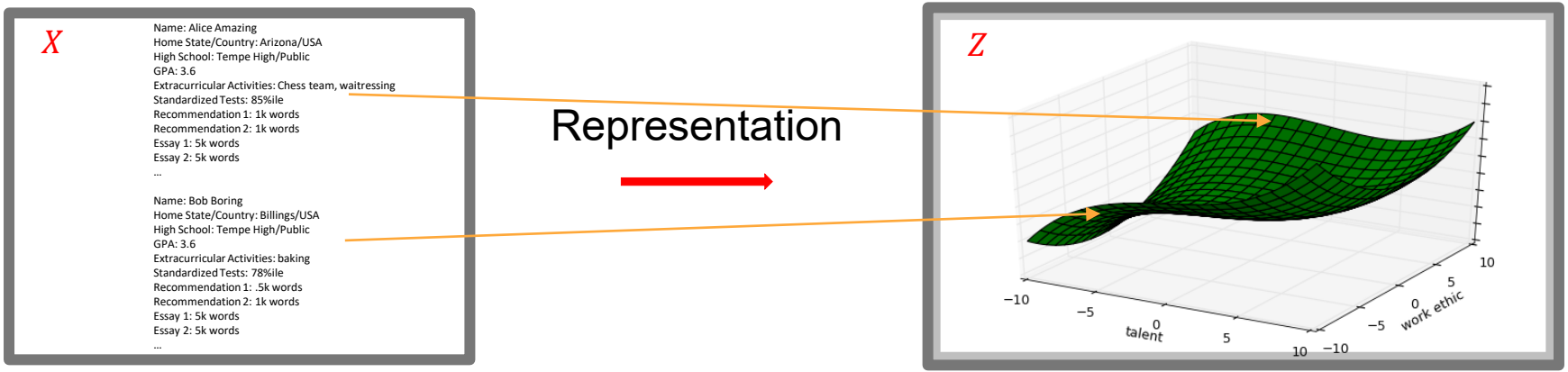
Locations considered in Study 1

Locations considered in Study 2



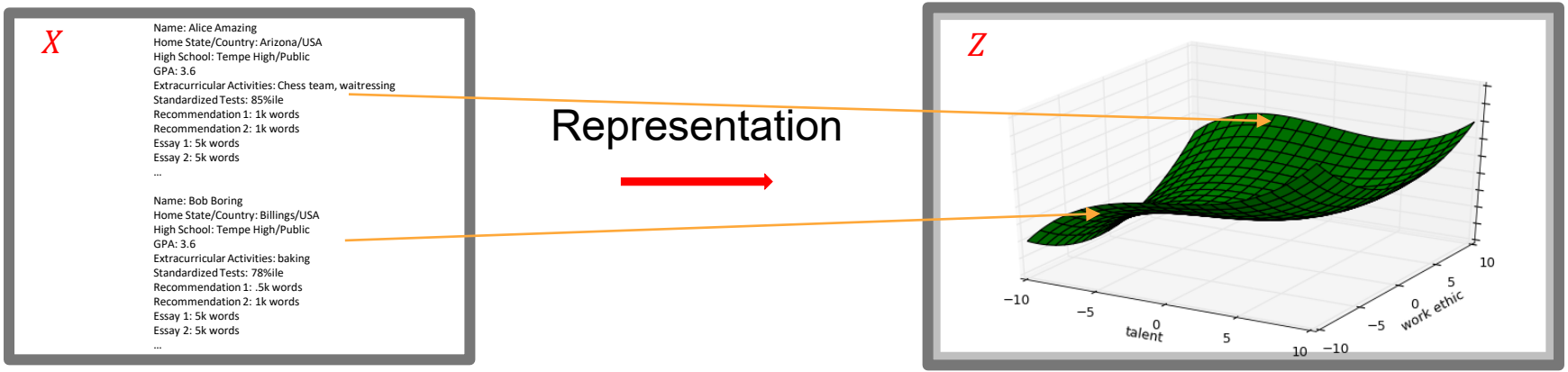
Representations (Informal)

- X: All possible real people
- Algorithm operates only on a representation of the person
  - The algorithm only knows what it is told about you
  - Distinct individuals may be mapped to the same representation



Representations (Informal)

- X: All possible real people
- Algorithm operates only on a representation of the person
  - The algorithm only knows what it is told about you
  - Distinct individuals may be mapped to the same representation
  - We make no such assumption



Why are Algorithms Unfair?

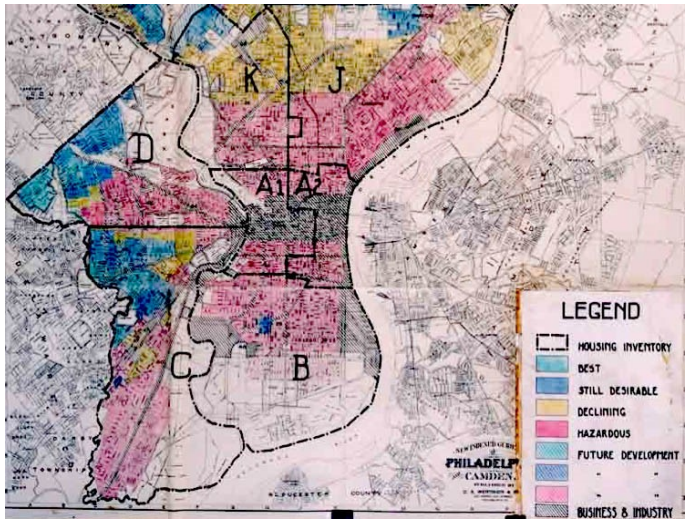
- Unrepresentative training data
- Training labels are historical decisions, which are biased
- Features used are differentially expressive
  - Zero AP classes
  - Access to Medicaid data but not to private medical insurance data
- Unbalanced outcome proxies
  - Referral bias and re-arrest bias instead of child abuse and recidivism

Paradigm

- Definitions: Define what it means for an algorithm to be fair
  - Catalog of evils: what do we wish the algorithm to *prevent*
  - Problematic for technical and social reasons
- Algorithms: Construct algorithms that are fair according to the definition
  - Some success
- Composition: Prove that systems built from fair pieces are fair *in toto*
  - Often not the case

(Incomplete) Catalog of Algorithmic Fairness Evils

- Explicit discrimination: explicitly test for membership in S and give less desirable outcome
- Redlining: Discrimination based on redundant encoding or property correlated with membership in S

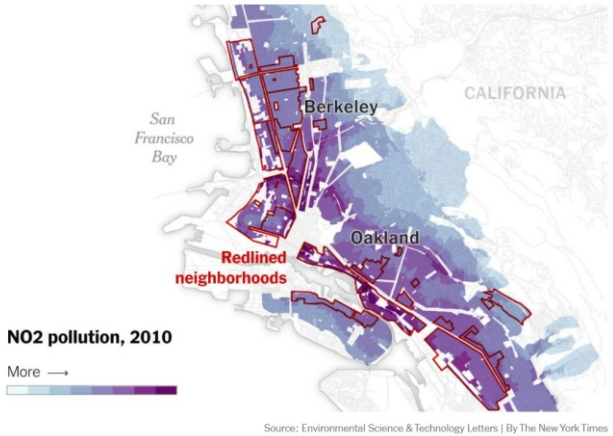


(Incomplete) Catalog of Algorithmic Fairness Evils

- Explicit discrimination: explicitly test for membership in S and give less desirable outcome
- Redlining: Discrimination based on redundant encoding or property correlated with membership in S

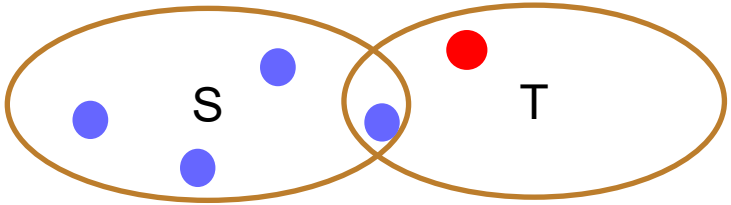
How Air Pollution Across America Reflects a Racist Policy From the 1930s

A new study shows how redlining, a Depression-era housing policy, contributed to inequalities that persist decades later in U.S. cities.



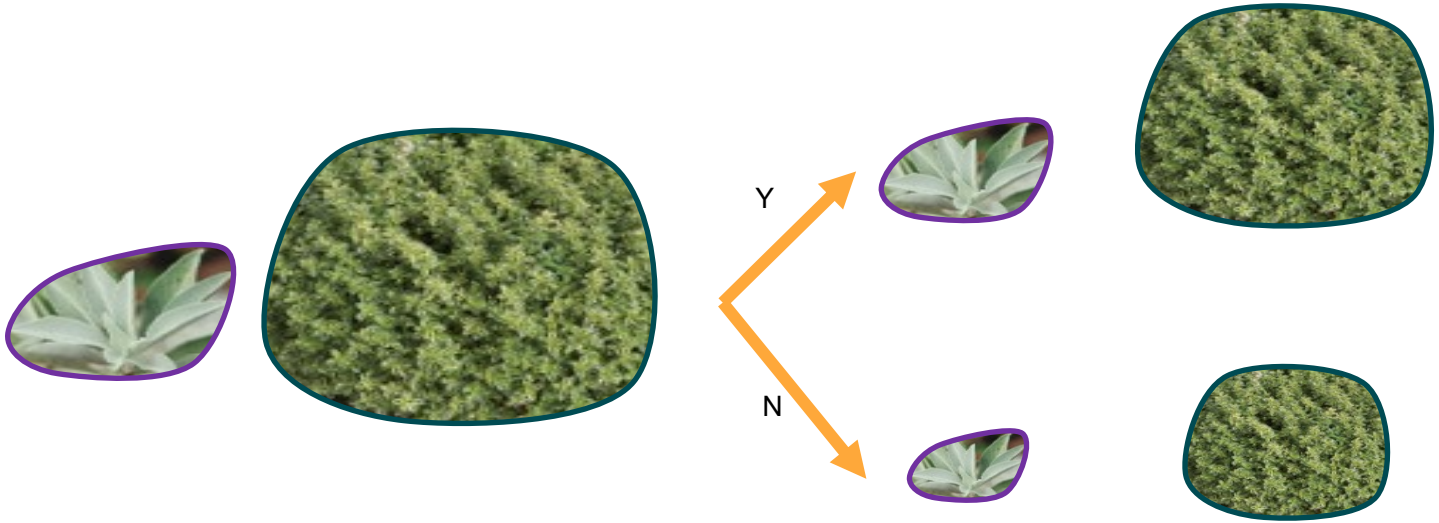
(Incomplete) Catalog of Algorithmic Fairness Evils

- Explicit discrimination: explicitly test for membership in S and give less desirable outcome
- Redlining: Discrimination based on redundant encoding or property correlated with membership in S
- Cutting off business with an S-heavy segment of the population
- Deliberately choosing "wrong" members of S (possibly in order to build a bad track record for S)
- Reverse tokenism: denying service to a highly qualified member of T – the token rejectee
- *Your evil here!*



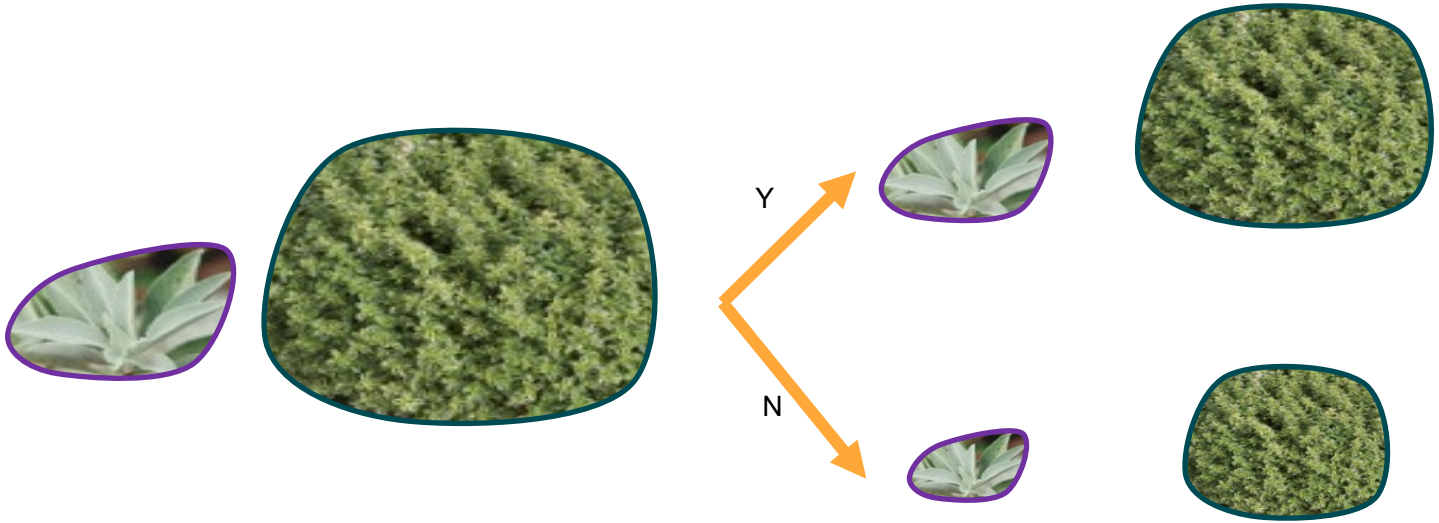
Defining Fairness for Groups

- Group fairness properties are statistical requirements
  - Statistical Parity: demographics of people assigned positive (negative) classification are the same as the demographics of the general population



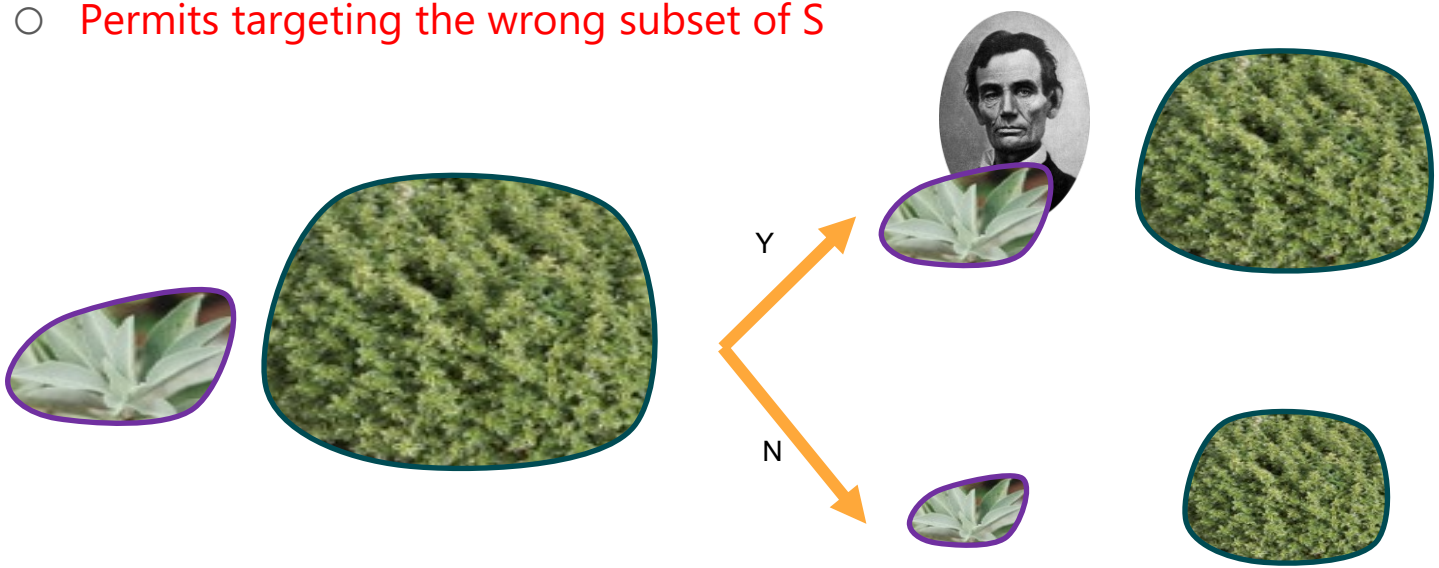
Defining Fairness for Groups

- Group fairness properties are statistical requirements
  - Statistical Parity: demographics of people assigned positive (negative) classification are the same as the demographics of the general population
    - May be meaningful in the breach; flawed as a solution concept



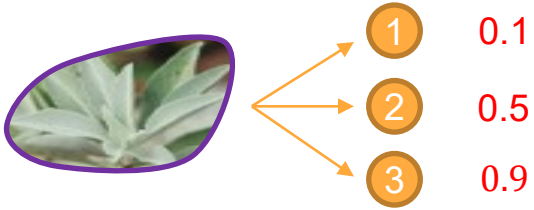
Defining Fairness for Groups

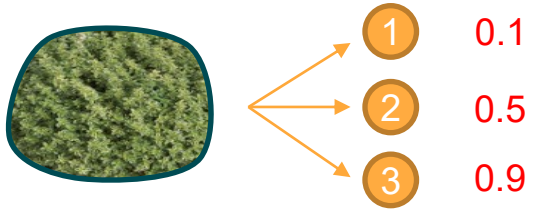
- Group fairness properties are statistical requirements
  - Statistical Parity: demographics of people assigned positive (negative) classification are the same as the demographics of the general population
  - Permits targeting the wrong subset of S



Defining Fairness for Groups

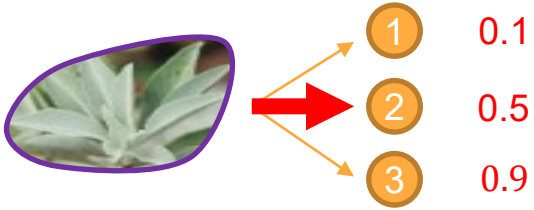
- Group fairness properties are statistical requirements
  - Calibration within groups: for each group $G \in \{S, T\}$ and each bin b with associated score v, "*v is correct on G in expectation*"

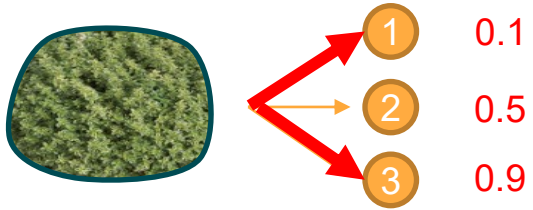




Defining Fairness for Groups

- Group fairness properties are statistical requirements
  - Calibration within groups: for each group $G \in \{S, T\}$ and each bin b with associated score v, "*v is correct on G in expectation*"
    - Nothing forces even remotely comparable differentiation
    - Example: both groups composed of equal numbers of 0.1's and 0.9's

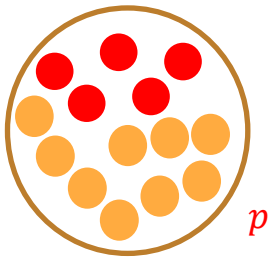




Unequal Error Rates as Group Unfairness

- Equal False Positive Rate (FPR) across groups
- Equal False Negative Rate (FNR) across groups
- Equal Positive Predictive Value (PPV) across groups

No *imperfect* classifier can simultaneously ensure equal FPR, FNR, PPV unless the base rates are equal



$p = 5/13$

$$\text{FPR} = \left(\frac{p}{1-p}\right)\left(\frac{1-\text{PPV}}{\text{PPV}}\right)(1-\text{FNR})$$

Chouldechova 2016

Proof

- Write

$$\mathrm{FPR_S} = \left(\frac{p_S}{1-p_S}\right)\left(\frac{1-\mathrm{PPV_S}}{\mathrm{PPV}_S}\right)(1-\mathrm{FNR_S})$$

$$\mathrm{FPR_T} = \left(\frac{p_T}{1-p_T}\right)\left(\frac{1-\mathrm{PPV_T}}{\mathrm{PPV}_T}\right)(1-\mathrm{FNR_T})$$

Chouldechova 2016

Pop Quiz

Is the problem solved by introducing a human into the process?

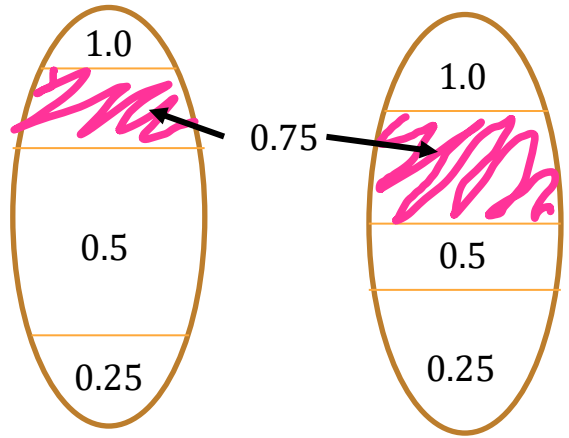
Pop Quiz

What about a classification deity?

Scoring Functions

- *A general group fairness goal*:
  $E_{x\sim D}[A(R(x))|x \in S, t(x) = v] =$
  $E_{x\sim D}[A(R(x))|x \in T, t(x) = v]$
  - Truly similar individuals receive, *on average*, similar treatment, independent of group
  - No promise about correctness

- *Compare to Calibration for Groups*
  - Implies that the scores "mean" the same thing across groups.



Half get 1, half get 0.2

All get 0.6

Inconsistency for Scoring Functions

- *Balance for positive class*: $E_{(x,y)}[f(x)|x \in S, y = 1] = E_{(x,y)}[f(x)|x \in T, y = 1]$
- *Balance for negative class*: $E_{(x,y)}[f(x)|x \in S, y = 0] = E_{(x,y)}[f(x)|x \in T, y = 0]$
- f is calibrated on both S and T

These three conditions cannot be satisfied simultaneously for imperfect predictors unless the base rates are the same in S and T.

Kleinberg, Mullainathan, Raghavan 2016

Surprisingly Difficult to Audit

Benchmark Test for Police Stops

- Examines differences of rates of police contact, scaled by race-specific rates of expected contact were the police not to discriminate

- Nature of the inference: higher scaled rate of contact indicates discrimination
- Problem: The choice of denominator strongly affects the conclusion

Neil and Winship, 2019

Kohler-Hausmann (2018) makes a particularly powerful case against definitions of discrimination that are based on the notion of individuals who are similarly situated but for race. She argues that if we think race is a social construct—as most social scientists do—then it does not make sense to speak of discrimination as the treatment effect of race, as race is not something confounded by other features but rather something that is constituted by those features. Put differently, were we capable of controlling for everything else, there would be no treatment—no solid-state race—that remained because those features and relations exist in a complex interrelationship that together constitute racial categories as we know them. She argues that whether or not something is discrimination is thus a normative question, which can only be made sense of with situated cultural knowledge about the relevant categories of stratification. As an example, she describes a hypothetical audit study in which male and female job candidates are sent out wearing the same dresses: They are not similarly situated candidates but for sex, because sex makes the dress mean something different.

Benchmark Test

Table 1 Hypothetical data and analyses

	S	T
Data		
Population	100,000	100,000
Criminals	15,000	10,000
Stops	10,000 (10%)	5,000 (5%)
Searches	5,000 (50%)	1,250 (25%)
Hits	250 (5%)	125 (10%)
Analyses		
Population-based benchmark test for stops	(10,000/100,000):(5,000/100,000) = 2:1	
Criminal-based benchmark test for stops	(10,000/15,000):(5,000/10,000) = 1.33:1	
Stop-based benchmark test for searches	(5,000/10,000):(1,250/5,000) = 2:1	
Outcome test for searches	(250/5,000):(125/1,250) = 1:2	

Benchmark Test

$$\frac{|Stops(S)|}{|S|} : \frac{|Stops(T)|}{|T|}$$

Table 1 Hypothetical data and analyses

	S	T
Data		
Population	100,000	100,000
Criminals	15,000	10,000
Stops	10,000 (10%)	5,000 (5%)
Searches	5,000 (50%)	1,250 (25%)
Hits	250 (5%)	125 (10%)
Analyses		
Population-based benchmark test for stops	(10,000/100,000):(5,000/100,000) = 2:1	
Criminal-based benchmark test for stops	(10,000/15,000):(5,000/10,000) = 1.33:1	
Stop-based benchmark test for searches	(5,000/10,000):(1,250/5,000) = 2:1	
Outcome test for searches	(250/5,000):(125/1,250) = 1:2	

Benchmark Test

$$\frac{|Stops(S)|}{|Crim(S)|} : \frac{|Stops(T)|}{|Crim(T)|}$$

Table 1 Hypothetical data and analyses

	S	T
Data		
Population	100,000	100,000
Criminals	15,000	10,000
Stops	10,000 (10%)	5,000 (5%)
Searches	5,000 (50%)	1,250 (25%)
Hits	250 (5%)	125 (10%)
Analyses		
Population-based benchmark test for stops	(10,000/100,000):(5,000/100,000) = 2:1	
Criminal-based benchmark test for stops	(10,000/15,000):(5,000/10,000) = 1.33:1	
Stop-based benchmark test for searches	(5,000/10,000):(1,250/5,000) = 2:1	
Outcome test for searches	(250/5,000):(125/1,250) = 1:2	

Extreme Example: Criminality in Denominator

- In S: 85,000 of 100K are innocent
- In T: 90 of 100K are innocent
- Stops(S)=Stops(T)
- ZERO criminals stopped in each group (epic police fail)
- Stops/innocent in S > Stops/innocent in T!

$$\frac{|Stops(S)|}{|Crim(S)|} : \frac{|Stops(T)|}{|Crim(T)|}$$

Criminality in Denominator is Problematic

Table 2 Using crime as a benchmark understates discrimination against the higher-crime group

	S		T	
Criminality	Innocent	Criminal	Innocent	Criminal
Number of individuals	85,000	15,000	90,000	10,000
Probability of stop	0.0647	0.30	0.0333	0.20
Stops	5,500	4,500	3,000	2,000
Criminal-based benchmark test for stops	(10,000/15,000):(5,000/10,000) = 1.33:1			

Police don't know who is a criminal. The data here are consistent with previous table. Probability of a stop for an innocent in S is nearly twice that for an innocent in T, so the 1.33 : 1 ratio underestimates the unfairness.

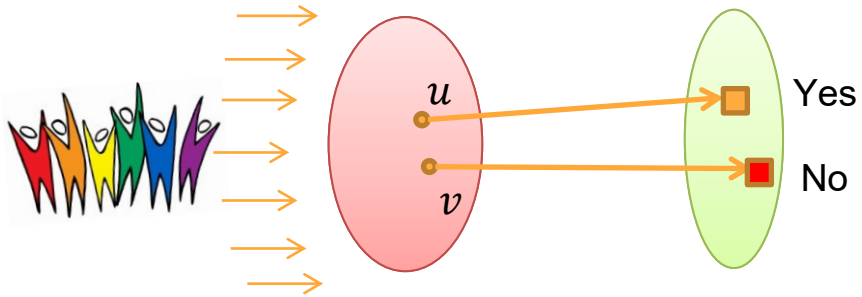
Suppose: police only stop people in public spaces, everyone in public space has probability 0.25 of being stopped, and 40,000 members of S, but only 20,000 members of T, use public spaces. “Similarly situated (but for S/T) are treated similarly;” no discrimination(??)

Table 1 Hypothetical data and analyses

	S	T
Data		
Population	100,000	100,000
Criminals	15,000	10,000
Stops	10,000 (10%)	5,000 (5%)
Searches	5,000 (50%)	1,250 (25%)
Hits	250 (5%)	125 (10%)
Analyses		
Population-based benchmark test for stops	(10,000/100,000):(5,000/100,000) = 2:1	
Criminal-based benchmark test for stops	(10,000/15,000):(5,000/10,000) = 1.33:1	
Stop-based benchmark test for searches	(5,000/10,000):(1,250/5,000) = 2:1	
Outcome test for searches	(250/5,000):(125/1,250) = 1:2	

Individual Fairness

"Similar people" are treated similarly

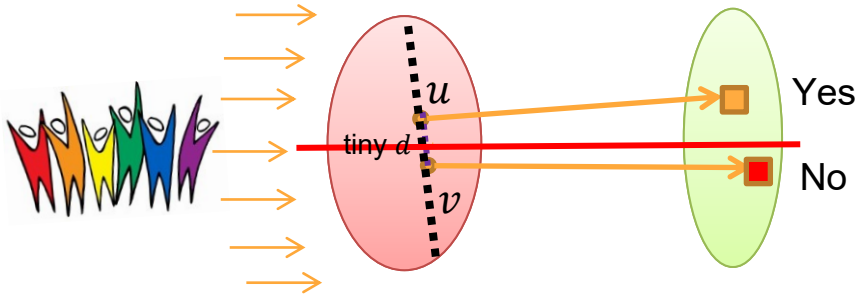


Need "right" notion of (dis)similarity $d(u, v)$ for the specific classification task

DworkHardtReingoldPitassiZemel 2012

Individual Fairness

"Similar people" are treated similarly

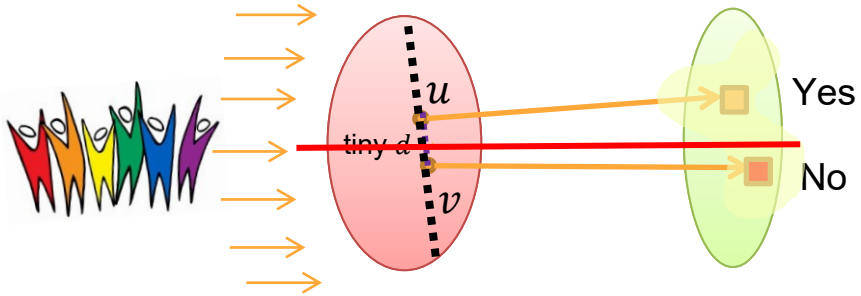


"You have to draw the line somewhere"

DworkHardtReingoldPitassiZemel 2012

Individual Fairness

"Similar people" have similar *probability distributions* on outcomes



"You have to draw the line somewhere"

Really?

DworkHardtReingoldPitassiZemel 2012

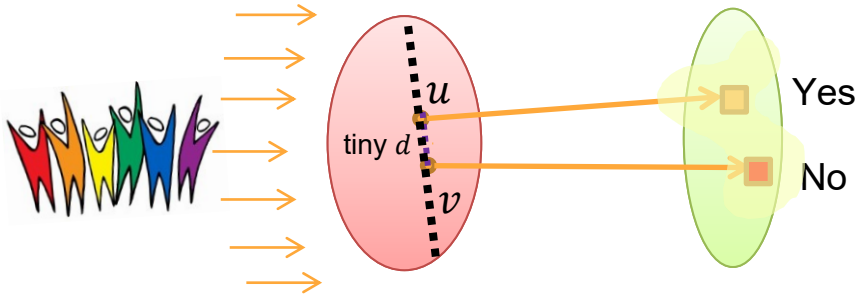
Algorithms for Individually Fair Classification

Perspective

- Fairness is a hard constraint, accuracy is best possible
- In contrast to the case in the traditional privacy notion in secure function evaluation (SFE, MPC): produce exact answers, privacy is best possible; and in contrast to maximizing revenue as a hard constraint with best possible fairness
- In this work, the implicit assumption was that the data are extremely rich, capturing everything needed for correct prediction, preparing for a dystopic future in which your computer – or your advertiser -- knows everything about you

Individual Fairness

"Similar people" have similar *probability distributions* on outcomes

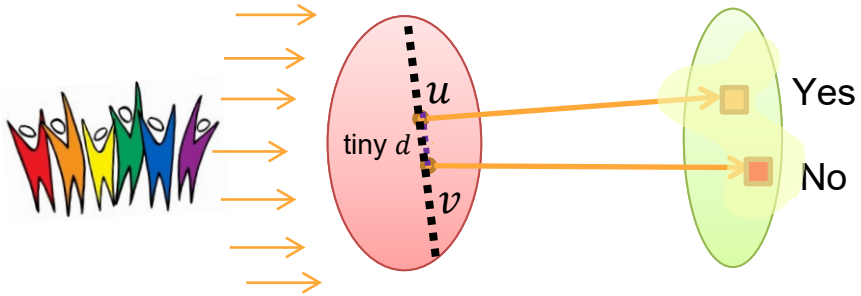


$$C: U \to \Delta(O)$$
$$||C(x) - C(y)|| \leq d(x, y)$$

$$D_{\text{tv}}(P, Q) = \frac{1}{2} \sum_{a \in A} |P(a) - Q(a)|$$

$$D_{\infty}(P, Q) = \sup_{a \in A} \log \left(\max \left\{ \frac{P(a)}{Q(a)}, \frac{Q(a)}{P(a)} \right\} \right)$$

Utility



$L(v, o)$ = "Loss" incurred by mapping v to o

Assemble Ingredients

"Minimize vendor's utility loss, subject to the fairness conditions."

Loss function: soft constraints

Fairness conditions: hard constraints

"Fairness LP"

$$\min_{\substack{M=\{\mu_u\}_{u\in U} \\ \mu_u \in \Delta(O)}} E_{u\in V} E_{o\sim\mu_u} L(u,o)$$

$$||\mu_u - \mu_v|| \leq d(u,v) \qquad \text{(Lipschitz)}$$

DworkHardtReingoldPitassiZemel 2012

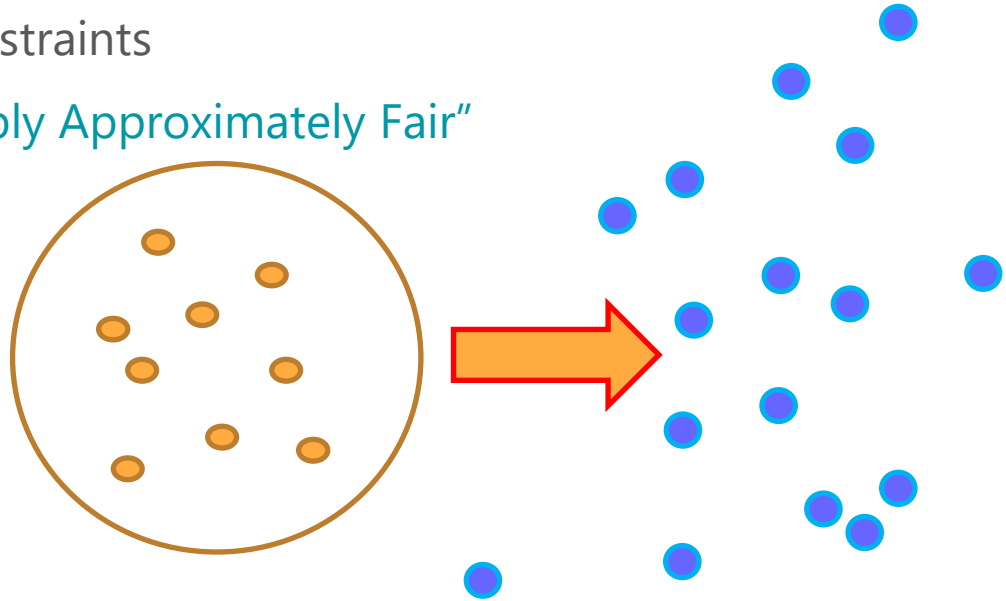
Assemble Ingredients

"Minimize utility loss, subject to the fairness conditions."

Loss function: soft constraints

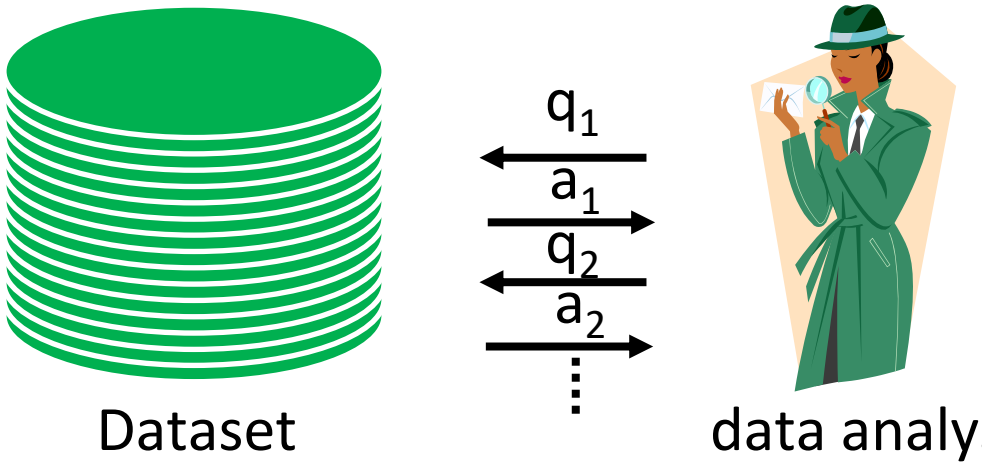
Fairness conditions: hard constraints

Generalizability via relaxing to "Probably Approximately Fair"



RothblumYona 2018

Privacy-Preserving Data Analysis



- Driving scenario: analysis of US Census data
- 55+ year old problem

Differential Privacy

M gives ε-differential privacy if for all pairs of adjacent data sets x, y, and all output events S

$$\Pr[\text{see } S \text{ on } M(x)] \leq e^{\varepsilon} \Pr[\text{see } S \text{ on } M(y)]$$

"Privacy Loss" bound

Randomness introduced by M

DworkMcSherryNissimSmith'06

The Exponential Mechanism

- $f(x) \in \Xi = \{\xi_1, \xi_2, \ldots, \xi_k\}$
  - Strings, experts, small databases, prices, models, …

    Each $\xi \in \Xi$ has a utility for x, denoted $u(x, \xi)$

    $\Delta(u) = \max_{\xi, \text{adj } x,y} |u(x, \xi) - u(y, \xi)|$

- Intuition: Output ξ with probability $\propto e^{u(x,\xi)\epsilon/\Delta u}$

$$\left[\frac{\exp\big(u(x,\xi)\big)}{\exp\big(u(y,\xi)\big)}\right]^{\epsilon/\Delta u} = \left[e^{u(x,\xi)-u(y,\xi)}\right]^{\epsilon/\Delta u} \leq e^{\epsilon}$$

The Exponential Mechanism

- $f(x) \in \Xi = \{\xi_1, \xi_2, \ldots, \xi_k\}$
  - Strings, experts, small databases, prices…

    Each $\xi \in \Xi$ has a utility for x, denoted $u(x, \xi)$

    $\Delta(u) = \max_{\xi, \text{adj } x,y} |u(x, \xi) - u(y, \xi)|$

- Formally, output ξ with probability $\frac{e^{u(x,\xi)\epsilon/2\Delta u}}{\sum_{\xi'} e^{u(x,\xi')\epsilon/2\Delta u}}$

Normalization term

- Proof of ϵ-DP: ratio of numerators $\leq \exp(\frac{\epsilon}{2})$;

  ratio of denominators $\geq \exp(-\frac{\epsilon}{2})$

Utility of Exponential Mechanism

- Theorem: Let $\Xi^* \subseteq \Xi$ be the set of optimal-utility outputs for x: $u(x, \xi^*) = OPT = u^* \; \forall \xi^* \in \Xi^*$.

  Then $\forall v$: $\Pr[u(x, M(x)) \leq v] \leq \frac{|\Xi| e^{\epsilon v / 2\Delta u}}{|\Xi^*| e^{\epsilon u^* / 2\Delta u}} = \frac{|\Xi|}{|\Xi^*|} e^{\epsilon (v - u^*)/2\Delta u}$

- Proof:
  - If $u(x, \xi) \leq v$ then un-normalized weight of ξ is at most $e^{\epsilon v / 2\Delta u}$. There are at most $|\Xi|$ such elements.
  - There are $|\Xi^*|$ elements with un-normalized weight $e^{\epsilon u^* / 2\Delta u}$ so the normalization term is at least $|\Xi^*| e^{\epsilon u^* / 2\Delta u}$

Using DP to Obtain Fairness: Preliminaries

Differential Privacy automatically ensures (degraded) privacy for groups

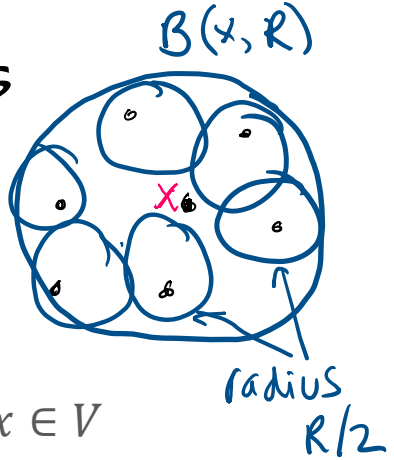
- If M is ε-DP then it is $k\varepsilon$-DP for groups of size k (homework)
- Generally, ε-differential privacy ensures that for all events E, and for all not necessarily adjacent x, y: $\Pr[E|x] \leq e^{\varepsilon \cdot |x \Delta y|} \Pr[E|y]$
  - $x \Delta y$ set of elements appearing in just one of x and y

Using DP to Obtain Fairness: Preliminaries

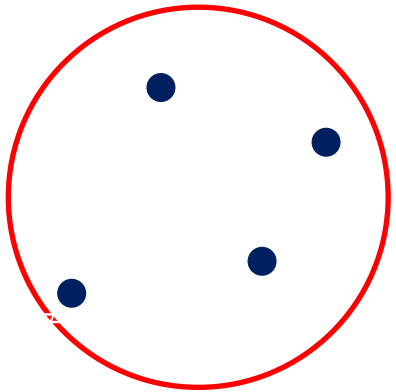
- $\mu_x \in \Delta(O)$ maps an individual x to a distribution on outcome space O
- A Lipschitz constraint on $d_\infty(\mu_x, \mu_y) \leq d(x, y)$ translates to familiar form:
  - $\mu_x(a) \leq e^{d(x,y)}\mu_y(a)$ for all $a \in O$ (but remember: x and y are individuals, not databases!)
  - For fixed x, y these are linear constraints
  - $\mu_x \in \Delta(O)$ is captured by $0 \leq \mu_x(a) \leq 1$ and $\sum_{a \in O} \mu_x(a) = 1$
- Let β_x be the distribution over individuals v $\in V$: $\beta_x(v) = \frac{e^{-d(x,v)}}{\sum_{v \in V} e^{-d(x,v)}}$
- Claim (homework): $\left|\left|\beta_x - \beta_y\right|\right|_\infty \leq 2d(x, y)$.

Denominator $\triangleq N_x \geq 1$ because $d(x, x) = 0$

Using DP to Obtain Fairness: Preliminaries



- Doubling Dimension of (V, d)
  - Least k such that $\forall x \in V, \forall R > 0$
    $B(x, R) = \{y \in V \mid d(x, y) \leq R\}$ can be covered by 2^k balls of radius $R/2$
- (V, d) is well-separated if $\exists \varepsilon > 0$ such that $|B(x, \varepsilon)| = 1$ for all $x \in V$
- Fact: Suppose S is a set of points in a metric space with doubling dimension $\leq k$. If
  S is contained in some ball of radius r and
  $\forall y, z \in S$ such that $y \neq z$, $d(y, z) > \varepsilon$,
  then $|S| \leq (4r/\varepsilon)^k$.



Using DP to Obtain Fairness

Theorem: Let (d, V) be well separated with bounded doubling dimension. Then $\{\beta_x\}_{x \in V}$ satisfies $E_{x \in V} E_{y \sim \beta_x} d(x, y) = O(1)$.

Proof: Fact $\Rightarrow |B(x, r)| \leq (4r/\varepsilon)^k = (4/\varepsilon)^k r^k = 2^{O(k)} r^k$.

$$E_{x \in V} E_{y \sim \beta_x} d(x, y) \leq E_{x \in V} E_{\substack{y \sim \beta_x \\ d(x,y) \leq 1}} d(x, y) + E_{x \in V} E_{\substack{y \sim \beta_x \\ d(x,y) \geq 1}} d(x, y)$$

$$\leq 1 + E_{x \in V} E_{\substack{y \sim \beta_x \\ d(x,y) \geq 1}} d(x, y)$$

$$\beta_x(v) = \frac{e^{-d(x,v)}}{\sum_{v \in V} e^{-d(x,v)}}$$

$$E_{x\in V}E_{y\sim\beta_x}d(x,y) \le 1 + E_{x\in V}E_{\substack{y\sim\beta_x \\ d(x,y)\ge 1}} d(x,y)$$

$$1 + E_{x\in V}E_{\substack{y\sim\beta_x \\ d(x,y)\ge 1}} d(x,y)$$

$$\le 1 + E_{x\in V}\int_1^\infty r\,\frac{e^{-r}}{N_x}|B(x,r)|dr$$

- $d(x,y)$
- $\beta_x(y)$ for y at distance r
- $\#y$ at distance $\le r$

$$\le 1 + E_{x\in V}\int_1^\infty r\,e^{-r}|B(x,r)|dr$$

because $N_x \ge 1$

$$\beta_x(v) = \frac{e^{-d(x,v)}}{\Sigma_{v\in V}e^{-d(x,v)}}$$

$$E_{x\in V}E_{y\sim\beta_x}d(x,y) \le 1 + E_{x\in V}\int_1^\infty r\,e^{-r}|B(x,r)|dr$$

$$1 + E_{x\in V}\int_1^\infty r\,e^{-r}|B(x,r)|dr = 1 + \int_1^\infty r\,e^{-r}E_{x\in V}|B(x,r)|dr$$

$$\le 1 + \int_1^\infty r\,e^{-r}E_{x\in V}\left(\frac{4r}{\varepsilon}\right)^k dr$$

$|B(x,r)| \le (4r/\varepsilon)^k$

$$\le 1 + \left(\frac{4}{\varepsilon}\right)^k \int_1^\infty r^{k+1}\,e^{-r}dr$$

$$\le 1 + \left(\frac{4}{\varepsilon}\right)^k \int_0^\infty r^{k+1}\,e^{-r}dr$$

$$= 1 + 2^{O(k)}\Gamma(k+2) = 1 + 2^{O(k)}(k+1)! = O(1)$$

$\Gamma(z) = \int_0^\infty x^{z-1}\,e^{-x}dx$

$\Gamma(n) = (n-1)!$ for positive integer n