Fairness and discrimination in actuarial pricing

Arthur Charpentier[1], Laurence Barry[2], Vincent Grari[3],
Lamprier Sylvain[3], Detyniecki Marcin[4]

[1] Univeristé du Québec à Montréal [2] Chaire Pari [3] Sorbonne Université [4] AXA

University of Illinois Urbana-Champaign, February 2022

Context: European Gender Directive (2012)

European Commission

English EN

Home > Press corner >

Available languages: English

Press release | 20 December 2012

EU rules on gender-neutral pricing in insurance industry enter into force

Commission Européenne

français FR

Accueil > Coin presse >

Langues disponibles: français

Communiqué de presse | 20 décembre 2012

La réglementation de l'UE sur la tarification unisexe en matière d'assurance entre en vigueur

source https://ec.europa.eu/commission/presscorner/detail/en/IP_12_1430

Agenda & keywords

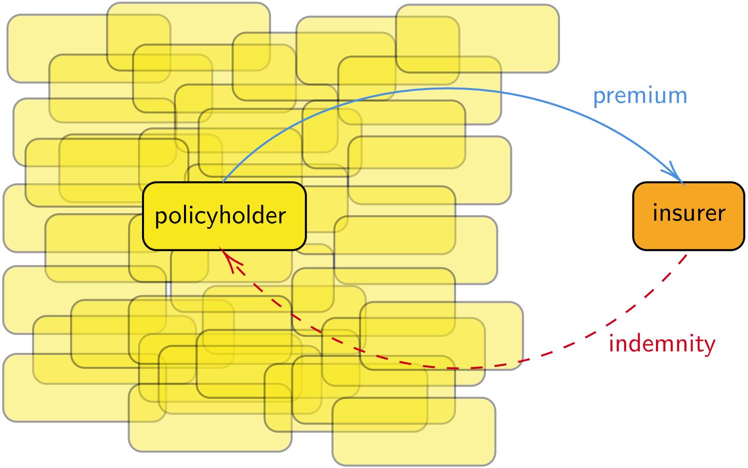
"*Technology is neither good nor bad; nor is it neutral*" , Kranzberg (1986)

- Insurance, mutualization, solidarity vs. individualization, heterogeneity
- Discrimination, *actuarial fairness*, legal aspects, discrimination by proxy
- Biases observation vs. experiment, selection bias, omited variable bias
- Fairness, $\widehat{Y} \perp\!\!\!\perp P$, $\widehat{Y} \perp\!\!\!\perp P \mid Y$ or $Y \perp\!\!\!\perp P \mid \widehat{Y}$, and individual fairness (counterfactual)
- Explainability and interpretability

see Charpentier (2022), Barry and Charpentier (2022) and Grari et al. (2022) for further details

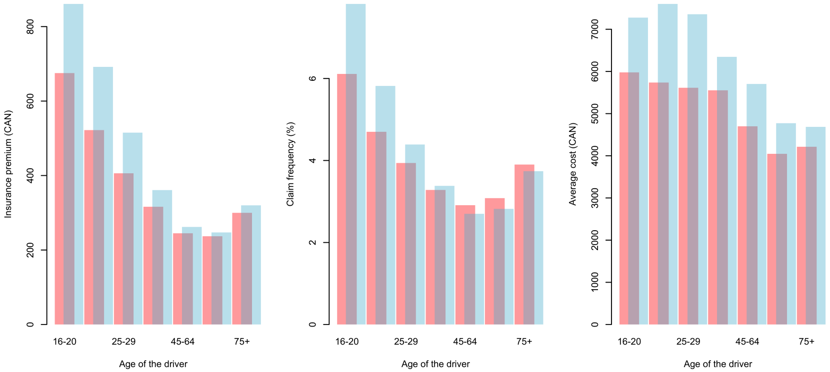
Insurance, risk pooling & solidarity

- Insurance is the contribution of the many to the misfortune of the few



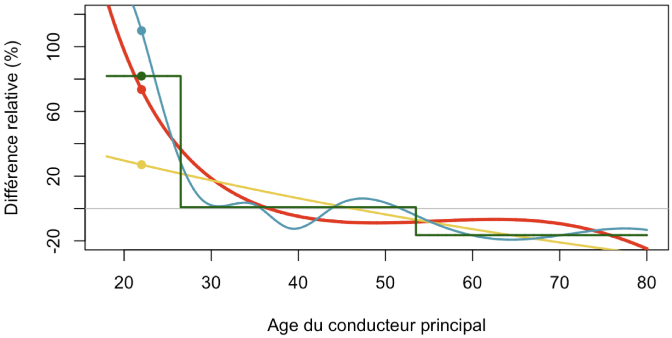
Heterogenous Risks I

- Insurance premium (CAN), claim frequency (%) and average cost (CAN) as a function of the age and the gender of the driver (?) in Québec



Heterogenous Risks II

- Claim frequency, as a function of the age of the driver (data Charpentier (2014))



- "*actuaries smoothed because smoothing was a 'mathematical and ethical' good*", Bouk (2015)
- interpretability and explanability, "*young drivers are more likely to have an accident*"

Heterogenous Risks III

- Life insurance, life tables function of the age and the genrder
- Men / women table, 1720 (Struyck (1912), page 231)

men						women					
0	1000	29.0%	45	371	16.6%	0	1000	28.9%	45	423	11.8%
5	710	5.6%	50	313	19.2%	5	711	5.2%	50	373	14.7%
10	670	4.2%	55	253	22.9%	10	674	3.3%	55	318	18.2%
15	642	5.5%	60	195	27.2%	15	652	4.3%	60	260	21.2%
20	607	6.6%	65	142	31.7%	20	624	5.8%	65	205	26.8%
25	567	7.9%	70	97	37.1%	25	588	6.8%	70	150	33.3%
30	522	9.2%	75	61	45.9%	30	548	7.3%	75	100	45.0%
35	474	10.5%	80	33	51.5%	35	508	7.9%	80	55	56.4%
40	424	12.5%	85	16		40	468	9.6%	85	24	

Heterogenous Risks IV

- Life insurance, life table as a function of the age and the gender
- More recent French life tables (TV, TD et INED)

TD 73-77	
0	100000
10	97961
20	97105
30	95559
40	93516
50	88380
60	77772
70	57981
80	28364
90	4986
100	103
110	0

TV 73-77	
0	100000
10	98447
20	98055
30	97439
40	96419
50	94056
60	89106
70	78659
80	52974
90	14743
100	531
110	0

TD 88-90	
0	100000
10	98835
20	98277
30	96759
40	94746
50	90778
60	81884
70	65649
80	39041
90	9389
100	263
110	0

TV 88-90	
0	100000
10	99129
20	98869
30	98371
40	97534
50	95752
60	92050
70	84440
80	65043
90	24739
100	1479
110	2

INED (M)	
0	100000
10	99486
20	99281
30	98656
40	97661
50	95497
60	90104
70	78947
80	59879
90	25123
100	1412

INED (F)	
0	100000
10	99578
20	99471
30	99247
40	98810
50	97645
60	94777
70	89145
80	77161
90	44236
100	4874

Heterogenous Risks V

- Life insurance, residual life expectancy (in years) as a function of the age, the gender and a smoker (or not) status, (data Benjamin and Michaelson (1988) 1970-1975, US)
- Hoffman (1931), Johnston (1945) "*it is clear that smoking is an important cause of mortality*", Miller and Gerstein (1983)

	men	
	non-smoker	smoker
25	48.4	42.8
35	38.7	33.3
45	29.2	24.2
55	20.3	16.5
65	12.8	10.4

	women	
	non-smoker	smoker
25	52.8	49.8
35	43.0	40.1
45	33.5	31.0
55	24.5	22.6
65	16.2	15.1

Heterogenous Risks VI

- Life insurance, life expectancy (in years) as a function of the age, the gender and the weight (BMI) (data Steensma et al. (2013) US)

  regular $[18.5; 25 kg/m^2)$, over-weighted $[25; 30 kg/m^2)$, obesity I $[30; 35 kg/m^2)$, obesity II $[35, 100 kg/m^2)$
- Crossley (2005), Czerniawski (2007) or Kelly and Markowitz (2009)

	men			
	regular	over.	obesity I	obesity II
20	57.2	61.0	59.1	53.5
30	47.6	51.4	49.4	44.1
40	38.1	41.7	39.9	34.7
50	28.9	32.4	30.6	25.8
60	20.4	23.6	21.9	17.6
70	13.2	15.8	14.4	10.9

	women			
	regular	over.	obesity I	obesity II
20	62.8	66.5	64.6	59.3
30	53.0	56.7	54.8	49.5
40	43.3	46.9	45.0	39.9
50	33.8	37.3	35.5	30.6
60	24.9	28.1	26.4	21.9
70	16.8	19.7	18.2	14.3

Heterogenous Risks VII

- handicap and genetic testing
- "*the insurance industry has generally regarded handicapped persons as undesirable risks*" Baker and Karol (1977)
- "*the denial of insurance coverage to an individual whose (non-inherited) cancer had been long cured would not constitute genetic discrimination, while the denial of insurance to that individual's relatives because of the (erroneous) belief that that type of cancer is heritable would be genetic discrimination*" Natowicz et al. (1992)
- Schatz (1986), Clifford and Iuculano (1987) (HIV), Jacobs and Sommers (2015) (inference from drug prescriptions)

Insurance and premium "individualization" I

- "*It is important to distinguish two things when talking about insurance. The first, the insurance operation, is technical and has a collective dimension, the second, the insurance contract, is legal and has an individual dimension*", Bigot and Cayol (2020) (aussi Thiery and Van Schoubroeck (2006), Lehtonen and Liukko (2015))
- **Individualistic approach**
  - The individualistic approach to equality analyses fundamental rights, such as the right to equal treatment, in terms of individuals.
  - An individual cannot be treated differently because of his or her membership in such a group, particularly in a group to which he or she has not chosen to belong.
- **Group approach**
  - The insurance tradition, on the other hand, analyses risks, premiums and benefit schedules in terms of groups
  - Unlike the individualistic approach, insurance classification schemes rely on the assumption that individuals answer to the average (stereotypical) characteristics of a group to which they belong.

Insurance and premium "individualization" II

- "*at the core of insurance business lies discrimination between risky and non-risky insureds*", Avraham (2017), see also Austin (1983) ("*insurance classification controversy*"), Frezal and Barry (2019), Barry (2020)
- perfect segmentation with observable latent risk factor Θ

	policyholder	insurer
loss	$\mathbb{E}[Y\vert\Theta]$	$Y-\mathbb{E}[Y\vert\Theta]$
average loss	$\mathbb{E}[Y]$	0
variance	$\text{Var}\Big[\mathbb{E}[Y\vert\Theta]\Big]$	$\text{Var}\Big[Y-\mathbb{E}[Y\vert\Theta]\Big]$

$$\text{Var}[Y] = \underbrace{\mathbb{E}\Big[\text{Var}[Y|\Theta]\Big]}_{\to\text{ insurer}} + \underbrace{\text{Var}\Big[\mathbb{E}[Y|\Theta]\Big]}_{\to\text{ policyholder}}.$$

Insurance and premium "individualization" III

- statistical segmentation with observable features $\boldsymbol{X} = (X_1, \cdots, X_k)$
  "*categorization based on immutable characteristics*", Crocker and Snow (2013)

	policyholder	insurer
loss	$\mathbb{E}[Y\vert\boldsymbol{X}]$	$Y - \mathbb{E}[Y\vert\boldsymbol{X}]$
average loss	$\mathbb{E}[Y]$	0
variance	$\text{Var}\Big[\mathbb{E}[Y\vert\boldsymbol{X}]\Big]$	$\mathbb{E}\Big[\text{Var}[Y\vert\boldsymbol{X}]\Big]$

$$\begin{aligned}\mathbb{E}\Big[\text{Var}[Y|\boldsymbol{X}]\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\text{Var}[Y|\Theta]\Big|\boldsymbol{X}\Big]\Big] + \mathbb{E}\Big[\text{Var}\Big[\mathbb{E}[Y|\Theta]\Big|\boldsymbol{X}\Big]\Big] \\ &= \underbrace{\mathbb{E}\Big[\text{Var}[Y|\Theta]\Big]}_{\text{perfect segmentation}} + \underbrace{\mathbb{E}\Big\{\text{Var}\Big[\mathbb{E}[Y|\Theta]\Big|\boldsymbol{X}\Big]\Big\}}_{\text{misfit}}.\end{aligned}$$

- "*kanssolidariteit*" vs "*subsidierende solidariteit*", De Pril and Dhaene (1996)

Assurance(s) & solidarité

- insurance health
- collective insurance
- natural catastrophes

  "*La Nation proclame la solidarité et l'égalité de tous les Français devant les charges qui résultent des calamités nationales*", Constitution du 27 octobre 1946

  "*solidarity in insurance means deciding not to segment the corresponding risk market on the basis of observable characteristics of individuals' risks*", Gollier (2002).
- non-life insurance

  "*Tout ce qui n'est pas défendu par la Loi ne peut être empêché*", Déclaration des Droits de l'Homme et du Citoyen, 1789, art. 5

"*access to insurance means not only the ability to purchase a policy for coverage, but perhaps also at an economically reasonable, non-prohibitive, non-discouraging cost*", Noguéro (2010)

Legal Aspects I

	CA	HI	GA	NC	NY	MA	PA	FL	TX	AL	ON	NB	NL	QC
Gender	×	×	●	×	●	×	×	●	●	●	●	×	×	●
Age	×	×	●	×*	●	×	●	●	●	●*	●	×	×	●
Driving experience	●	×	●	●	●	●	●	●	●	●	●	●	●	●
Credit history	×	×	●	●	●	×	●*	●	●	×*	×	●*	×	●
Education	×	×	×	×	×	×	●	●	●	●	●	●	●	●
Occupation	×	×	×	●	×	×	●	●	●	●	●	●	●	●
Employment status	×	×	×	●	×	×	●	●	●	●	●	●	●	●
Marital status	●	×	●	●	●	×	●	●	●	●	●	●	●	●
Housing situation	×	×	●	●	●	×	●	●	●	×	×	●	●	●
Address/ZIP code	●	●	●	●	●	●	●	●	●	×	×	●	●	●
Insurance history	●	●	●	●	●	●	●	●	●	●	●	●	●	●

CA: Californie, HI: Hawaii, GA: Georgia, NC: Caroline du nord, NY: New York, MA: Massachusetts, PA: Pennsylvanie, FL: Floride, TX: Texas

Bureau d'Assurance du Canada (2021)

Legal Aspects II

En France,

- le sexe ou le genre (art. A. 111-6 du Code des assurances, Commission européenne (Arr. 18 déc. 2012, NOR : EFIT1238658A, relatif à l'égalité entre les hommes et les femmes en assurance, JO 20 déc., mod. par Arr. 3 févr. 2014, NOR : EFIT1400411A, JO 11 févr.)
- distinction fondée sur l'âge (C. pén., art. 225-1 et 225-2), ("*belonging to a particular race or sex is akin to joining one specific 'club' at the moment of conception, whereas age...*", Macnicol (2006))
- la situation de famille ou sur l' orientation sexuelle (C. pén., art. 225-1 et 225-2)
- en raison du lieu de résidence d'une personne constitue une discrimination au sens pénal (C. pén., art. 225-1)
- "*Nul ne peut faire l'objet de discriminations en raison de ses* caractéristiques génétiques" (C. C., art. 16-13)

Legal Aspects III

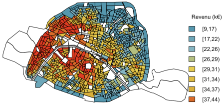
au Québec

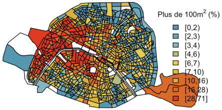
- “Toute personne a droit à la reconnaissance et à l’exercice, en pleine égalité, des droits et libertés de la personne, sans distinction, exclusion ou préférence fondée sur la race, la couleur, le sexe, l’identité ou l’expression de genre, la grossesse, l’orientation sexuelle, l’état civil, l’âge sauf dans la mesure prévue par la loi, la religion, les convictions politiques, la langue, l’origine ethnique ou nationale, la condition sociale, le handicap ou l’utilisation d’un moyen pour pallier ce handicap.” (C-12 - Charte des droits et libertés de la personne, art. 10)
- “*la distinction fondée sur l’âge, le sexe ou l’état civil est permise lorsqu’elle repose sur un facteur qui permet de déterminer un risque. Par exemple, une compagnie d’assurance peut vous poser des questions sur votre âge et votre sexe pour fixer votre prime*” (art. 20.1)

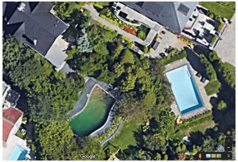
Proxy Based Discrimination (?) I

- location (policyholder home adress)

  Jean et al. (2016), Seresinhe et al. (2017), Gebru et al. (2017), Law et al. (2019), Ilic et al. (2019), Kita and Kidziński (2019), see also redlining







Proxy Based Discrimination (?) II

- location (policyholder home adress)

  Jean et al. (2016), Seresinhe et al. (2017), Gebru et al. (2017), Law et al. (2019), Ilic et al. (2019), Kita and Kidziński (2019), see also redlining

Proxy Based Discrimination (?) III

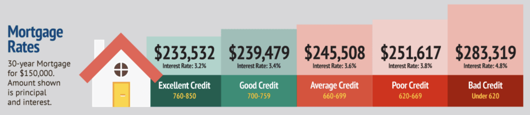
- facial recognition, recently some insurers have considered the idea of using facial recognition to predict certain diseases, Shikhare (2021)

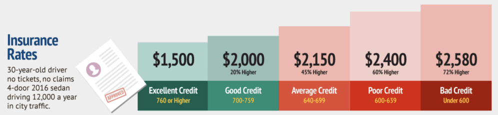
source https://nvlabs-fi-cdn.nvidia.com/stylegan2-ada-pytorch/, cf Karras et al. (2020)

- cf "phrénologie" Lombroso (1876) et Bertillon and Chervin (1909)
- cf "ugly laws" TenBroek (1966) et Burgdorf and Burgdorf Jr (1974)

Proxy Based Discrimination (?) IV

- credit scoring,
  Kabler (2004), Arya et al. (2013), Miller et al. (2003) Bartik and Nelson (2016), O'Neil (2016), Lauer (2017), Morris et al. (2017), Kiviat (2019)

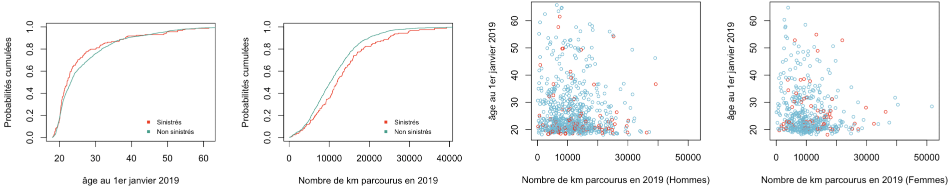




source https://www.incharge.org/debt-relief/credit-counseling/

Proxy Based Discrimination (?) V

- telematics, “behavioral" approach



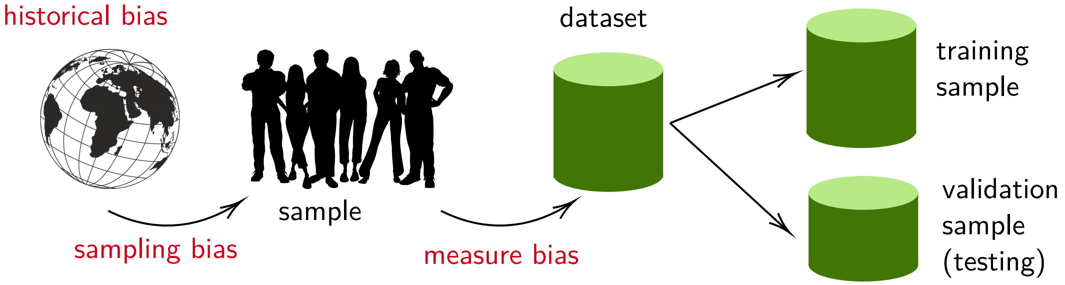
E.g. driven distance and policyholder gender Verbelen et al. (2018)

to go further on discrimination

- Notion of sensitive variable in GDPR
- Strong cultural component
- In high dimension (many explanatory variables $\boldsymbol{x}$), there are strong chances to have variables (strongly) correlated with a sensitive variable

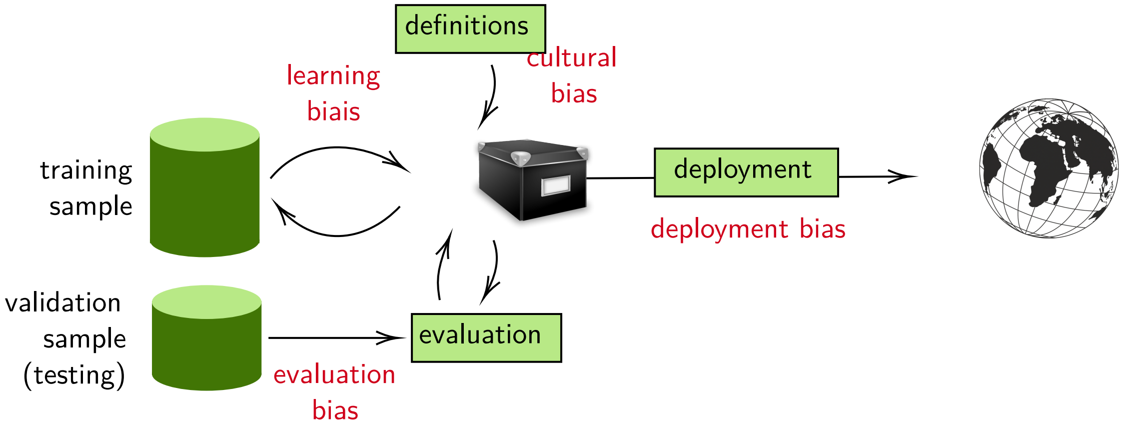
	Total	Men	Women	Proportions
Total	5233/12763 ∼ 41%	3714/8442 ∼ **44%**	1512/4321 ∼ 35%	66%-34%
Top 6	1745/4526 ∼ 39%	1198/2691 ∼ **45%**	557/1835 ∼ 30%	59%-41%
A	597/933 ∼ 64%	512/825 ∼ 62%	89/108 ∼ **82%**	88%-12%
B	369/585 ∼ 63%	353/560 ∼ 63%	17/ 25 ∼ **68%**	96%- 4%
C	321/918 ∼ 35%	120/325 ∼ **37%**	202/593 ∼ 34%	35%-65%
D	269/792 ∼ 34%	138/417 ∼ 33%	131/375 ∼ **35%**	53%-47%
E	146/584 ∼ 25%	53/191 ∼ **28%**	94/393 ∼ 24%	33%-67%
F	43/714 ∼ 6%	22/373 ∼ 6%	24/341 ∼ **7%**	52%-48%

Biases in Data Generation



(inspired by Suresh and Guttag (2019)).

Biases in Modeling



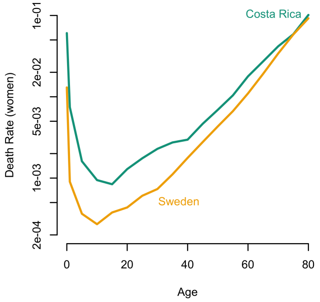
(inspired by Suresh and Guttag (2019)).

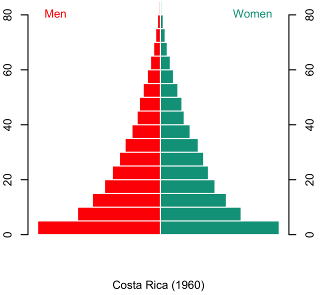
Simpsons Paradox & Eecological Fallacy

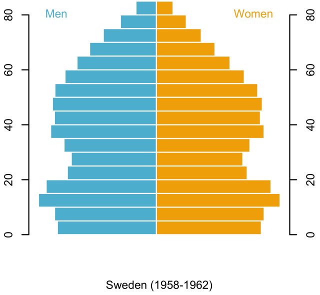
- Simpson's paradox, ecological fallacy⋆ (missing important variables)

  ex: number of accidents pedestrian-cars and maximum speed, Davis (2004)

  ex: mortality rate comparisons (local vs. global) Cohen (1986)







Retroaction & Goodhart Law

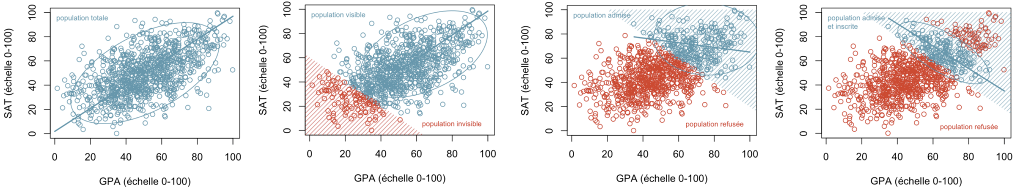
- Goodhart's law (and feedback bias)

  "*when a measure becomes a goal, it ceases to be a good measure*"

  ex: telematics and gamification

  "*insurers can eliminate uncertainty by shaping behavior*", Jarvis et al. (2019)

  ex: covid-related data, Giles (2020)



To go further on biases

- "*dark data*" by Hand (2020)
- Data We Know Are Missing
- Data We Dont Know Are Missing
- Choosing Just Some Cases
- Self-Selection
- Missing What Matters
- Data Which Might Have Been
- Changes with Time

- Definitions of Data
- Summaries of Data
- Measurement Error
- Feedback and Gaming
- Information Asymmetry
- Intentionally Darkened Data
- Fabricated and Synthetic Data
- Extrapolating beyond Your Data

Measuring and quantifying equity I

Notations:

$$\begin{cases} y \in \{0,1\} & \text{variable of interest} \\ p \in \{0,1\} & \text{protected variable (sensitive)} \\ \boldsymbol{x} \in \mathbb{R}^d & \text{`explanatory' variables} \\ s \in [0,1] & \text{score, classically } s = s(\boldsymbol{x}, p) \\ \widehat{y} \in \{0,1\} & \text{classifier, classically } \widehat{y} = \mathbf{1}(s > t) \end{cases}$$

Fairness Through Unawareness, Kusner et al. (2017)
Protected attribute p is not explicitly used in decision function $\widehat{y}$.

Measuring and quantifying equity II

Demographic Parity, (Corbett-Davies et al. (2017), Agarwal (2021))
Decision function $\widehat{y}$ satisfies demographic parity if $\widehat{Y} \perp\!\!\!\perp P$, i.e.

$$\mathbb{P}[\widehat{Y} = y | P = 0] = \mathbb{P}[\widehat{Y} = y | P = 1], \ \forall y \text{ or } \mathbb{E}[\widehat{Y} | P = 0] = \mathbb{E}[\widehat{Y} | P = 1]$$

In practice, compare $DI(\widehat{y}, p)$ (disparate impact) with 80%

$$DI(\widehat{Y}, P) = \frac{\mathbb{P}[\widehat{Y} = 1 | P = 0]}{\mathbb{P}[\widehat{Y} = 1 | P = 1]} \overset{?}{\leq} 80\%$$

see Feldman et al. (2015), Mercat-Bruns (2016) ou Biddle (2017) used by the State of California Fair Employment Practice Commission (FEPC) since 1971
(see also Besse et al. (2021))

Measuring and quantifying equity III

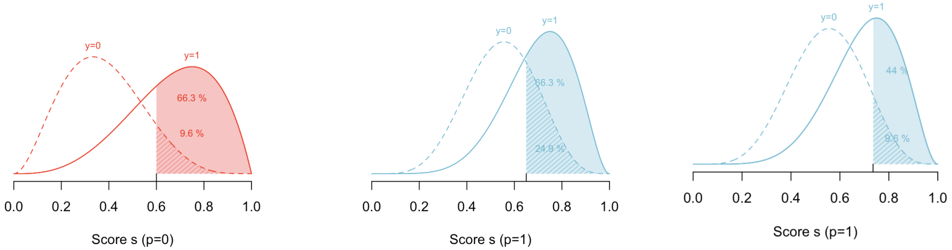
Equal Opportunity, Hardt et al. (2016)

True positive parity

$$\mathbb{P}[\widehat{Y}=1|P=0,Y=1]=\mathbb{P}[\widehat{Y}=1|P=1,Y=1]$$

or false positive parity

$$\mathbb{P}[\widehat{Y}=1|P=0,Y=0]=\mathbb{P}[\widehat{Y}=1|P=1,Y=0]$$



Measuring and quantifying equity IV

Equalized Odds, Hardt et al. (2016)
The parity of false positives and true positives is called equal opportunity,

$$\begin{cases}\mathbb{P}[\widehat{Y}=1|P=0,Y=1]=\mathbb{P}[\widehat{Y}=1|P=1,Y=1]\\ \mathbb{P}[\widehat{Y}=1|P=0,Y=0]=\mathbb{P}[\widehat{Y}=1|P=1,Y=0]\end{cases}$$

or

$$\mathbb{P}[\widehat{Y}=1|P=0,Y=y]=\mathbb{P}[\widehat{Y}=1|P=1,Y=y],\ \forall y\in\{0,1\}$$

i.e., $\widehat{Y}\perp\!\!\!\perp P$ conditionnal on Y.

Etc... there are many concepts, often incompatible with each other.

Measuring and quantifying equity V

statistical parity	Dwork et al. (2012)	$\mathbb{P}[\widehat{Y}=1\vert P=p]=\text{cst},\ \forall p$	independence
conditional statistical parity	Corbett-Davies et al. (2017)	$\mathbb{P}[\widehat{Y}=1\vert P=p,X=x]=\text{cst}_x,\ \forall p,y$	$\widehat{Y}\perp\!\!\!\perp P$
equalized odds	Hardt et al. (2016)	$\mathbb{P}[\widehat{Y}=1\vert P=p,Y=y]=\text{cst}_y,\ \forall p,y$	separation
equalized opportunity	Hardt et al. (2016)	$\mathbb{P}[\widehat{Y}=1\vert P=p,Y=1]=\text{cst},\ \forall p$	
predictive equality	Corbett-Davies et al. (2017)	$\mathbb{P}[\widehat{Y}=1\vert P=p,Y=0]=\text{cst},\ \forall p$	$\widehat{Y}\perp\!\!\!\perp P\mid Y$
balance (positive)	Kleinberg et al. (2017)	$\mathbb{E}[S\vert P=p,Y=1]=\text{cst},\ \forall p$	$S\perp\!\!\!\perp P\mid Y$
balance (negative)	Kleinberg et al. (2017)	$\mathbb{E}[S\vert P=p,Y=0]=\text{cst},\ \forall p$	
conditional accuracy equality	Berk et al. (2017)	$\mathbb{P}[Y=y\vert P=p,\widehat{Y}=y]=\text{cst}_y,\ \forall p,y$	sufficiency
predictive parity	Chouldechova (2017)	$\mathbb{P}[Y=1\vert P=p,\widehat{Y}=1]=\text{cst},\ \forall p$	
calibration	Chouldechova (2017)	$\mathbb{P}[Y=1\vert P=p,S=s]=\text{cst}_s,\ \forall p,s$	$Y\perp\!\!\!\perp P\mid\widehat{Y}$
well-calibration	Chouldechova (2017)	$\mathbb{P}[Y=1\vert P=p,S=s]=s,\ \forall p,s$	
accuracy equality	Berk et al. (2017)	$\mathbb{P}[\widehat{Y}=Y\vert P=p]=\text{cst},\ \forall p$	
treatment equality	Berk et al. (2017)	$\dfrac{\text{FN}_p}{\text{FP}_p}=\text{cst}_p,\ \forall p$	

Measuring and quantifying equity VI

Lipschitz property, Duivesteijn and Feelders (2008)

$$D(\widehat{y}_i,\widehat{y}_j) \text{ ou } D(s_i,s_j) \leq d(\boldsymbol{x}_i,\boldsymbol{x}_j),\ \forall i,j=1,\cdots,n.$$

Cf formal intervention "$\boldsymbol{X}$ is fixed at $\boldsymbol{x}$", see "$do(\boldsymbol{X}=\boldsymbol{x})$" in Pearl (1998) (or simply $do(\boldsymbol{x})$), (historically, from Wright (1921), Neyman et al. (1923) or Rubin (1974) Holland (1986))

Conterfactual fairness, Kusner et al. (2017) If the prediction in the real world is the same as the prediction in the counterfactual world where the individual would have belonged to a different demographic group, we have counterfactual equity, i.e.

$$\mathbb{P}\big[Y^\star_{P\leftarrow p}=y|\boldsymbol{X}=\boldsymbol{x},P=p\big]=\mathbb{P}\big[Y^\star_{P\leftarrow p'}=y|\boldsymbol{X}=\boldsymbol{x},P=p\big],\ \forall p',\boldsymbol{x},y.$$

To go further on quantifying fairness

LE JOURNAL DE MONTRÉAL FÉVRIER 2022 21H30 -13°C

LE GUIDE DE L'AUTO

Un homme change de sexe pour faire baisser la facture de son assurance auto

FRÉDÉRIC MERCIER
Vendredi, 27 juillet 2018 13:42
MISE À JOUR Vendredi, 27 juillet 2018 13:42

Désirant faire baisser le montant de sa prime d'assurance, un automobiliste de l'Alberta a fait changer son sexe sur son certificat de naissance.

«J'ai profité d'une faille dans le système», a expliqué l'albertain de 24 ans en entrevue avec CBC News.

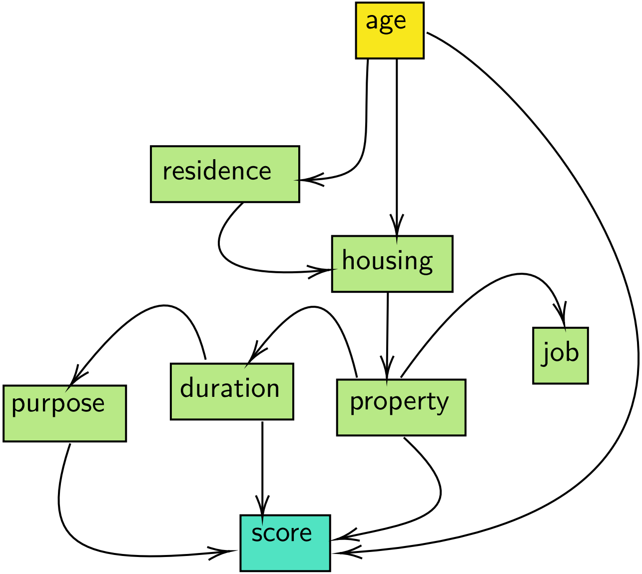
S'il n'a aucunement l'intention d'entreprendre de réelles procédures pour devenir une femme, le jeune homme désirant garder l'anonymat a tout de même fait changer son statut pour devenir «officiellement» une femme auprès du gouvernement albertain. Et il l'a fait uniquement pour économiser sur sa prime d'assurance.

Une différence marquée

L'idée de changement de sexe est venue au jeune homme après avoir appelé une compagnie d'assurance pour une soumission sur une voiture qu'il désirait acheter. Montant de la prime: 4517$ par année.

Curieux, le jeune homme a demandé à l'assureur combien lui coûterait une assurance sur le même véhicule s'il était une femme. On lui aurait alors répondu que la prime chuterait à 3423$ par année.

- P must be collected
- Looking for counterfactual
- DAGs are important



Quantifying fairness in actuarial models ($y \notin \{0, 1\}$)

Hirschfeld (1935), Gebelein (1941) and Rényi (1959)

$$HGR(U, V) = \max\left\{\text{corr}[f(U), g(V)]\right\} = \max_{f \in \mathcal{S}_U, g \in \mathcal{S}_V} \left\{\mathbb{E}[f(U)g(V)]\right\}$$

where $\mathcal{S}_U = \{f : \mathcal{U} \to \mathbb{R} : \mathbb{E}[f(U)] = 0 \text{ and } \mathbb{E}[f(U)^2] = 1\}$ and similarly $\mathcal{S}_V$.
One can also consider a conditional version,

$$HGR(U, V|Z) = \max_{f \in \mathcal{S}_{U|Z}, g \in \mathcal{S}_{V|Z}} \left\{\mathbb{E}[f(U)g(V)|Z]\right\}$$

where $\mathcal{S}_{U|Z} = \{f : \mathcal{U} \to \mathbb{R} : \mathbb{E}[f(U)|Z] = 0 \text{ and } \mathbb{E}[f(U)^2|Z] = 1\}$.

$$\begin{cases} \textbf{Demographic Parity}: & \widehat{Y} \perp\!\!\!\perp P & \text{i.e. } HGR(\widehat{Y}, P) = 0 \\ \textbf{Equalized Odds}: & \widehat{Y} \perp\!\!\!\perp P|Y & \text{i.e. } HGR(\widehat{Y}, P|Y) = 0 \end{cases}$$

Integrating fairness in a pricing model I

HGR can be difficult to estimate, but one can use some Neural Net,

$$HGR_{NN}(U,V)=\max_{\omega_f,\omega_g}\{\mathbb{E}[f_{\omega_f}(U)g_{\omega_g}(V)]\}$$

In a classical ML or econometric pricing model, solve

$$\underset{\boldsymbol{\theta}}{\text{argmin}}\left\{\mathcal{L}\big(\widehat{y},y\big)\right\},\text{ where }\mathcal{L}\big(\widehat{y},y\big)=\sum_{i=1}^{n}\ell\big(\widehat{y}_i,y_i\big)\text{ and }\widehat{y}=h_{\boldsymbol{\theta}}(x)$$

To avoid over-fit: penalize complexity (penalty $\mathcal{P}$)

$$\underset{\boldsymbol{\theta}}{\text{argmin}}\left\{\mathcal{L}\big(h_{\boldsymbol{\theta}}(\boldsymbol{x}),y\big)+\lambda\mathcal{P}\big(h_{\boldsymbol{\theta}}\big)\right\}$$

Integrating fairness in a pricing model II

Inspired by Goodfellow et al. (2018), to avoid un-fairness: penalize according to $HGR(\widehat{y},p)$ (for demographic parity)

$$\underset{\boldsymbol{\theta},\boldsymbol{\omega}_f,\boldsymbol{\omega}_g}{\text{argmin}}\left\{\mathcal{L}\big(h_{\boldsymbol{\theta}}(x),y\big)+\lambda \text{HGR}_{\boldsymbol{\omega}_f,\boldsymbol{\omega}_g}(\widehat{y},p)\right\}$$

i.e.

$$\underset{\boldsymbol{\theta}}{\text{argmin}}\left\{\max_{\boldsymbol{\omega}_f,\boldsymbol{\omega}_g}\left\{\mathcal{L}(h_{\boldsymbol{\theta}}(\boldsymbol{X}),Y)+\lambda\mathbb{E}_{(\boldsymbol{X},S)\sim\mathcal{D}}(\widehat{f}_{\boldsymbol{\omega}_f}(h_{\boldsymbol{\theta}}(\boldsymbol{X}))\widehat{g}_{\boldsymbol{\omega}_g}(P)\right\}\right\}$$

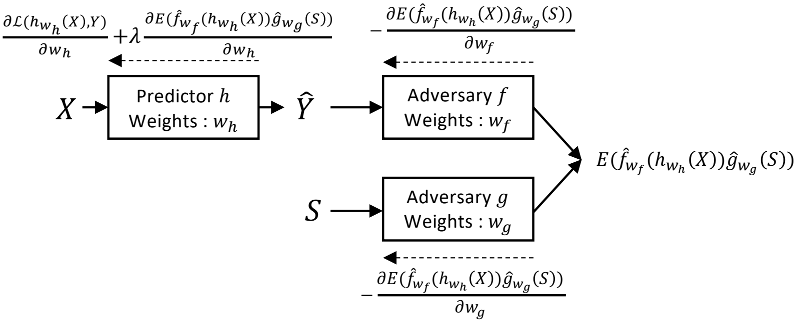
or $HGR(\widehat{y},p|y)$ (for equalized odds), i.e. when $y\in\{0,1\}$

$$\begin{aligned}\underset{\boldsymbol{\theta}}{\text{argmin}}\Bigg\{\max_{\boldsymbol{\omega}_{f0},\boldsymbol{\omega}_{g0},\boldsymbol{\omega}_{f1},\boldsymbol{\omega}_{g1}}\Big\{\mathcal{L}(h_{\boldsymbol{\theta}}(\boldsymbol{X}),Y)\quad&+\quad\lambda_0\mathbb{E}_{(\boldsymbol{X},P)\sim\mathcal{D}_0}(\widehat{f}_{\boldsymbol{\omega}_{f_0}}(h_{\boldsymbol{\theta}}(\boldsymbol{X}))\widehat{g}_{\boldsymbol{\omega}_{g_0}}(P))\\&+\quad\lambda_1\mathbb{E}_{(\boldsymbol{X},P)\sim\mathcal{D}_1}(\widehat{f}_{\boldsymbol{\omega}_{f_1}}(h_{\boldsymbol{\theta}}(\boldsymbol{X}))\widehat{g}_{\boldsymbol{\omega}_{g_1}}(P))\Big\}\Bigg\}\end{aligned}$$

Integrating fairness in a pricing model III

or, more generally when $y \in \Omega_Y$ (e.g. $\{0,1,2,3+\}$), if $k = \#\Omega_y$

$$\underset{\boldsymbol{\theta}}{\text{argmin}}\left\{\max_{\omega_{f0},\omega_{g0},\omega_{fk},\omega_{gk}}\left\{\mathcal{L}(h_{\boldsymbol{\theta}}(\boldsymbol{X}),Y)+\sum_{y\in\Omega_y}\lambda_y\mathbb{E}_{(\boldsymbol{X},P)\sim\mathcal{D}_y}\left(\widehat{f}_{\omega_{f_y}}(h_{\boldsymbol{\theta}}(\boldsymbol{X}))\widehat{g}_{\omega_{g_y}}(P)\right)\right\}\right\}$$



Dealing with high dimension I

- geographic / spatial information, $\boldsymbol{X}_g$
- car type / make / model, $\boldsymbol{X}_g$
- other classical ratemaking variables, $\boldsymbol{X}_p$ (non protected)

Some features can be in high dimension, natural solution would be PCA or autoencoders (see Shi and Shi (2021) about feature embedding in high dimension)

Dealing with high dimension II

Principal Component Analysis (PCA)

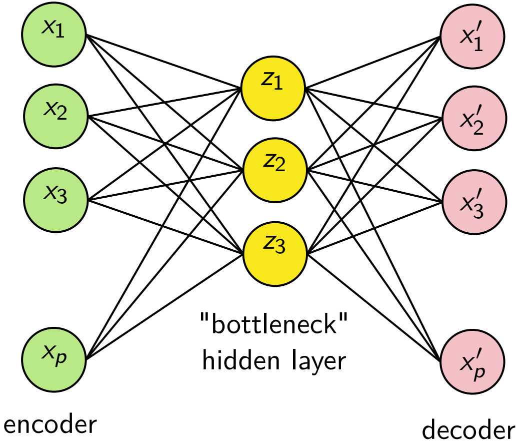
$$\min_{\boldsymbol{P}\in\Pi}\left\{\|\boldsymbol{X}-\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{X}\|_F^2\right\}\text{ s.t. }\operatorname{rank}(\boldsymbol{P})=k$$

where Π is the set of projection matrices.

$$\begin{tikzcd}
\boxed{\boldsymbol{X}\in\mathcal{X}} \arrow[r, bend right, "\boldsymbol{P}"'] & \boxed{\boldsymbol{Z}\in\mathcal{Z}} \arrow[r, bend right, "\boldsymbol{P}^\top"'] & \boxed{\boldsymbol{X}'\in\mathcal{X}}
\end{tikzcd}$$

$$\min\|\boldsymbol{X}-\boldsymbol{X}'\|^2=\min\|\boldsymbol{X}-\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{X}\|^2$$

$$=\min\sum_{i=1}^n\left(\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{x}_i-\boldsymbol{x}_i\right)^\top\left(\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{x}_i-\boldsymbol{x}_i\right)$$



Dealing with high dimension III

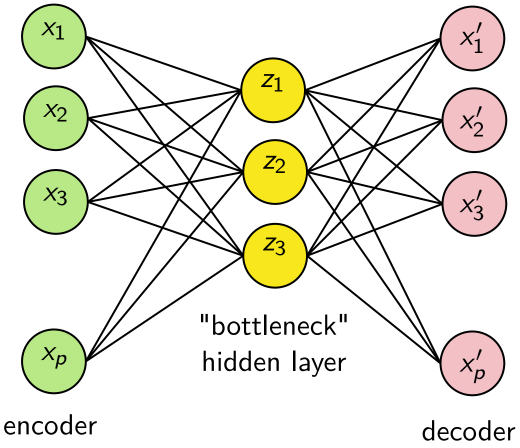
Autoencoder

$$\min_{\psi}\left\{\|\boldsymbol{X}-\psi\circ\varphi\boldsymbol{X}\|_F^2\right\}$$

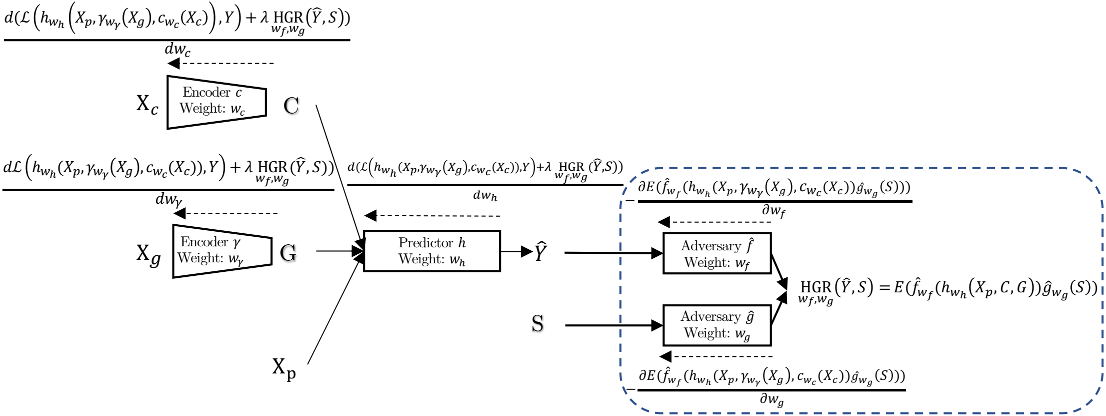
$$\boldsymbol{X}\in\mathcal{X} \xrightarrow{\varphi} \boldsymbol{Z}\in\mathcal{Z} \xrightarrow{\psi} \boldsymbol{X}'\in\mathcal{X}$$

$$\min\|\boldsymbol{X}-\boldsymbol{X}'\|^2=\min\|\boldsymbol{X}-\psi\circ\varphi(\boldsymbol{X})\|^2$$

$$\min\sum_{i=1}^{n}\big(\psi\circ\varphi(\boldsymbol{x}_i)-\boldsymbol{x}_i\big)^\top\big(\psi\circ\varphi(\boldsymbol{x}_i)-\boldsymbol{x}_i\big)$$

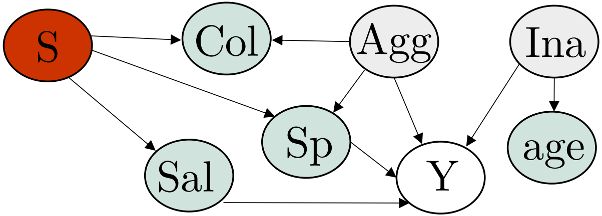


Dealing with high dimension IV

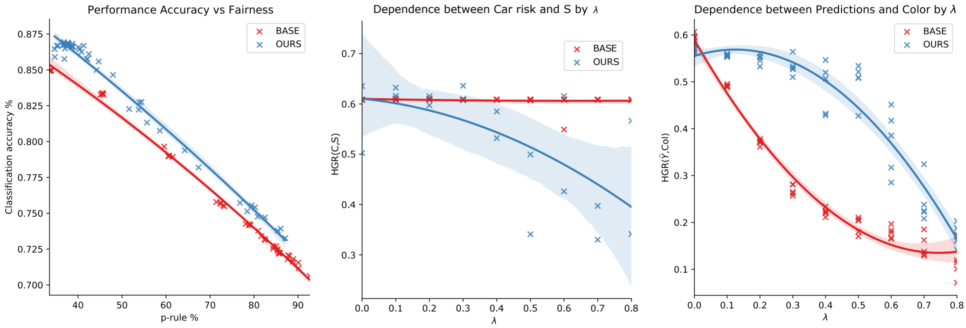


Application on synthetic data I

- S: sensitive / protected (gender)
- Col: color of the car
- Sp: maximum speed of the car
- Sal: average salary of the policyholders area
- Age: age of the driver
- Ina: inattention
- Agg: aggressivity
- Y: total cost



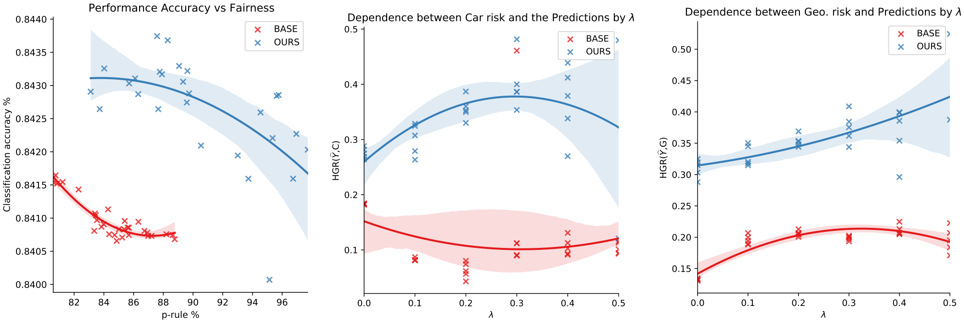
Application on synthetic data II



- λ: fairness penalty
- p-rule: $\min\left\{\dfrac{\mathbb{P}(\widehat{Y}=1|S=1)}{\mathbb{P}(\widehat{Y}=1|S=0)},\dfrac{\mathbb{P}(\widehat{Y}=1|S=0)}{\mathbb{P}(\widehat{Y}=1|S=1)}\right\}$

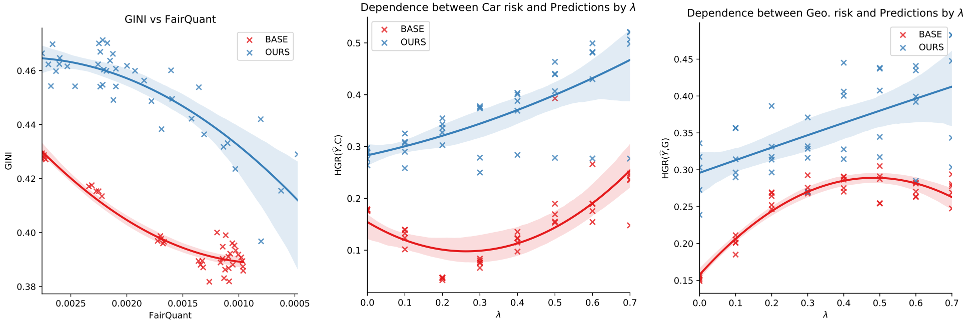
Application on real data (pricing game 2015) I

$y \in \{0, 1\}$ (claim occurrence)



Application on real data (pricing game 2015) II

$y \in \{0, 1, 2+\}$ (claim frequency)



see Grari et al. (2022) for more examples (including the case where $y \in \mathbb{R}^+$

References I

Agarwal, S. (2021). Trade-offs between fairness and interpretability in machine learning. In *IJCAI 2021 Workshop on AI for Social Good*.

Arya, S., Eckel, C., and Wichman, C. (2013). Anatomy of the credit score. *Journal of Economic Behavior & Organization*, 95:175–185.

Austin, R. (1983). The insurance classification controversy. *University of Pennsylvania Law Review*, 131(3):517–583.

Avraham, R. (2017). Discrimination and insurance. In Lippert-Rasmussen, K., editor, *Handbook of the Ethics of Discrimination*, pages 335–347. Routledge.

Baker, L. D. and Karol, C. (1977). Employee insurance benefit plans and discrimination on the basis of handicap. *DePaul L. Rev.*, 27:1013.

Barry, L. (2020). Insurance, big data and changing conceptions of fairness. *European Journal of Sociology*, 61:159 – 184.

Barry, L. and Charpentier, A. (2020). Personalization as a promise: Can big data change the practice of insurance? *Big Data & Society*, 7(1):2053951720935143.

Barry, L. and Charpentier, A. (2022). The Fairness of Machine Learning in Insurance: New Rags for an Old Man? . *ArXiv*.

References II

Bartik, A. and Nelson, S. (2016). Deleting a signal: Evidence from pre-employment credit checks. *SSRN*, 2759560.

Benjamin, B. and Michaelson, R. (1988). Mortality differences between smokers and non-smokers. *Journal of the Institute of Actuaries*, 115(3):519525.

Berk, R., Heidari, H., Jabbari, S., Joseph, M., Kearns, M., Morgenstern, J., Neel, S., and Roth, A. (2017). A convex framework for fair regression. *arXiv*, 1706.02409.

Bertillon, A. and Chervin, A. (1909). *Anthropologie métrique: conseils pratiques aux missionnaires scientifiques sur la manière de mesurer, de photographier et de décrire des sujets vivants et des pièces anatomiques*. Imprimerie nationale.

Besse, P., del Barrio, E., Gordaliza, P., Loubes, J.-M., and Risser, L. (2021). A survey of bias in machine learning through the prism of statistical parity. *The American Statistician*, 0(0):1–11.

Biddle, D. (2017). *Adverse impact and test validation: A practitioner's guide to valid and defensible employment testing*. Routledge.

Bigot, R. and Cayol, A. (2020). *Le droit des assurances en tableaux*. Ellipses.

Bigot, R. and Charpentier, A. (2019). Repenser la responsabilité, et la causalité. *Risques*, 120:123–128.

Bigot, R. and Charpentier, A. (2020). Quelle responsabilité pour les algorithmes? *Risques*, 121.

References III

Bouk, D. (2015). *How Our Days Became Numbered: Risk and the Rise of the Statistical Individual*. The University of Chicago Press.

Bureau d'Assurance du Canada (2021). Facts of the property and casualty insurance industry in canada.

Burgdorf, M. P. and Burgdorf Jr, R. (1974). A history of unequal treatment: The qualifications of handicapped persons as a suspect class under the equal protection clause. *Santa Clara Lawyer*, 15:855.

Charpentier, A. (2014). *Computational Actuarial Science*. The R series. CRC Press.

Charpentier, A. (2022). *Assurance: biais, discrimination et équité*. Institut Louis Bachelier.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2):153–163.

Clifford, K. A. and Iuculano, R. P. (1987). Aids and insurance: the rationale for aids-related testing. *Harvard law review*, 100(7):1806–1825.

Cohen, J. E. (1986). An uncertainty principle in demography and the unisex issue. *The American Statistician*, 40(1):32–39.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., and Huq, A. (2017). Algorithmic decision making and the cost of fairness. *arXiv*, 1701.08230.

References IV

Crocker, K. J. and Snow, A. (2013). The theory of risk classification. In Loubergé, H. and Dionne, G., editors, *Handbook of insurance*, pages 281–313. Springer.

Crossley, M. (2005). Discrimination against the unhealthy in health insurance. *U. Kan. L. Rev.*, 54:73–154.

Czerniawski, A. M. (2007). From average to ideal: The evolution of the height and weight table in the united states, 1836-1943. *Social Science History*, 31(2):273296.

Davis, G. A. (2004). Possible aggregation biases in road safety research and a mechanism approach to accident modeling. *Accident Analysis & Prevention*, 36(6):1119–1127.

De Pril, N. and Dhaene, J. (1996). Segmentering in verzekeringen. *DTEW Research Report 9648*, pages 1–56.

Duivesteijn, W. and Feelders, A. (2008). Nearest neighbour classification with monotonicity constraints. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 301–316. Springer.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226.

References V

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., and Venkatasubramanian, S. (2015). Certifying and removing disparate impact. In *proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 259–268.

Frezal, S. and Barry, L. (2019). Fairness in uncertainty: Some limits and misinterpretations of actuarial fairness. *Journal of Business Ethics*.

Gebelein, H. (1941). Das statistische problem der korrelation als variations- und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 21(6):364–379.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., and Fei-Fei, L. (2017). Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, 114(50):13108–13113.

Giles, C. (2020). Goodharts law comes back to haunt the uks covid strategy. *Financial Times*, 14-5.

Gollier, C. (2002). La solidarite sous langle economique. *Revue Générale du Droit des Assurances*, pages 824–830.

Goodfellow, I., McDaniel, P., and Papernot, N. (2018). Making machine learning robust against adversarial inputs. *Communications of the ACM*, 61(7):56–66.

References VI

Grari, V., Charpentier, A., Lamprier, S., and Detyniecki, M. (2022). A fair pricing model via adversarial learning. *ArXiv*, 2202.12008.

Hand, D. J. (2020). *Dark Data: Why What You Dont Know Matters*. Princeton University Press.

Hardt, M., Price, E., and Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29:3315–3323.

Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(4):520524.

Hoffman, F. L. (1931). Cancer and smoking habits. *Annals of surgery*, 93(1):50.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

Ilic, L., Sawada, M., and Zarzelli, A. (2019). Deep mapping gentrification in a large canadian city using deep learning and google street view. *PloS one*, 14(3):e0212814.

Jacobs, D. B. and Sommers, B. D. (2015). Using drugs to discriminateadverse selection in the insurance marketplace. *New England Journal of Medicine*.

Jarvis, B., Pearlman, R. F., Walsh, S. M., Schantz, D. A., Gertz, S., and Hale-Pletka, A. M. (2019). Insurance rate optimization through driver behavior monitoring. *Google Patents*, 10,169,822.

References VII

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794.

Johnston, L. (1945). Effects of tobacco smoking on health. *British Medical Journal*, 2(4411):98.

Kabler, B. (2004). Insurance-based credit scores: Impact on minority and low income populations in missouri. *State of Missouri Departement of Insurance*.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119.

Kelly, I. R. and Markowitz, S. (2009). Incentives in obesity and health insurance. *Inquiry*, 46(4):418–432.

Kita, K. and Kidziński, Ł. (2019). Google street view image of a house predicts car accident risk of its resident. *arXiv*, 1904.05270.

Kiviat, B. (2019). The moral limits of predictive practices: The case of credit-based insurance scores. *American Sociological Review*, 84(6):1134–1158.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, 133(1):237–293.

References VIII

Kranzberg, M. (1986). Technology and history:" kranzberg's laws". *Technology and culture*, 27(3):544–560.

Kusner, M. J., Loftus, J., Russell, C., and Silva, R. (2017). Counterfactual fairness. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 4066–4076. NIPS.

Lauer, J. (2017). *Creditworthy: A History of Consumer Surveillance and Financial Identity in America*. Columbia University Press.

Law, S., Paige, B., and Russell, C. (2019). Take a look around: Using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology*, 10(5).

Lehtonen, T.-K. and Liukko, J. (2015). Producing solidarity, inequality and exclusion through insurance. *Res Publica*, 21(2):155–169.

Lombroso, C. (1876). *L'uomo delinquente*. Hoepli.

Macnicol, J. (2006). *Age discrimination: An historical and contemporary analysis*. Cambridge University Press.

Mercat-Bruns, M. (2016). *Discrimination at Work*. University of California Press.

Miller, G. and Gerstein, D. R. (1983). The life expectancy of nonsmoking men and women. *Public Health Reports*, 98(4):343.

References IX

Miller, M. J., Smith, R. A., and Southwood, K. N. (2003). The relationship of credit-based insurance scores to private passenger automobile insurance loss propensity. *Actuarial Study, Epic Actuaries*.

Morris, D. S., Schwarcz, D., and Teitelbaum, J. C. (2017). Do credit-based insurance scores proxy for income in predicting auto claim risk? *Journal of Empirical Legal Studies*, 14(2):397–423.

Natowicz, M. R., Alper, J. K., and Alper, J. S. (1992). Genetic discrimination and the law. *American Journal of Human Genetics*, 50(3):465.

Neyman, J., Dabrowska, D. M., and Speed, T. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, pages 465–472.

Noguéro, D. (2010). Sélection des risques. discrimination, assurance et protection des personnes vulnérables. *Revue générale du droit des assurances*, 3:633–663.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods & Research*, 27(2):226–284.

Rényi, A. (1959). On measures of dependence. *Acta mathematica hungarica*, 10(3-4):441–451.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

References X

Schatz, B. (1986). The aids insurance crisis: Underwriting or overreaching. *Harvard Law Review*, 100:1782.

Seresinhe, C. I., Preis, T., and Moat, H. S. (2017). Using deep learning to quantify the beauty of outdoor places. *Royal Society open science*, 4(7):170170.

Shi, P. and Shi, K. (2021). Nonlife insurance risk classification using categorical embedding. *SSRN*, 3777526.

Shikhare, S. (2021). Next generation ltc - life insurance underwriting using facial score model. In *Insurance Data Science conference*.

Steensma, C., Loukine, L., Orpana, H., Lo, E., Choi, B., Waters, C., and Martel, S. (2013). Comparing life expectancy and health-adjusted life expectancy by body mass index category in adult canadians: a descriptive study. *Population health metrics*, 11(1):1–12.

Struyck, N. (1912). *Les oeuvres de Nicolas Struyck (1687-1769): qui se rapportent au calcul des chances, à la statistique général, z la statistique des décès et aux rentes viagèter*. Société générale néerlandaise d'assurances sur la vie et de rentes viagères.

Suresh, H. and Guttag, J. V. (2019). A framework for understanding sources of harm throughout the machine learning life cycle. *arXiv*, 1901.10002.

References XI

TenBroek, J. (1966). The right to live in the world: The disabled in the law of torts. *Calif. L. Rev.*, 54:841.

Thiery, Y. and Van Schoubroeck, C. (2006). Fairness and equality in insurance classification. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 31(2):190–211.

Verbelen, R., Antonio, K., and Claeskens, G. (2018). Unravelling the predictive power of telematics data in car insurance pricing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(5):1275–1304.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20.