Discrimination-Free Insurance Pricing

Mario V. Wüthrich
RiskLab, ETH Zurich

RiskLab® Switzerland

"Deep Learning with Actuarial Applications in R"
Swiss Association of Actuaries SAA/SAV, Zurich
October 14/15, 2021

Programme SAV Block Course

- Refresher: Generalized Linear Models (THU 9:00-10:30)
- Feed-Forward Neural Networks (THU 13:00-15:00)
- Discrimination-Free Insurance Pricing (THU 17:15-17:45)

- LocalGLMnet (FRI 9:00-10:30)
- Convolutional Neural Networks (FRI 13:00-14:30)
- Wrap Up (FRI 16:00-16:30)

Contents: Discrimination-Free Insurance Pricing

- Direct discrimination
- Indirect discrimination
- Unawareness price
- Discrimination-free price

- **Direct Discrimination**

Best-Estimate Pricing

- Basic pricing setup is given by
  - ⋆ Y denotes the claim costs;
  - ⋆ $\boldsymbol{X}$ denotes non-discriminatory covariates;
  - ⋆ $\boldsymbol{D}$ denotes discriminatory covariates.

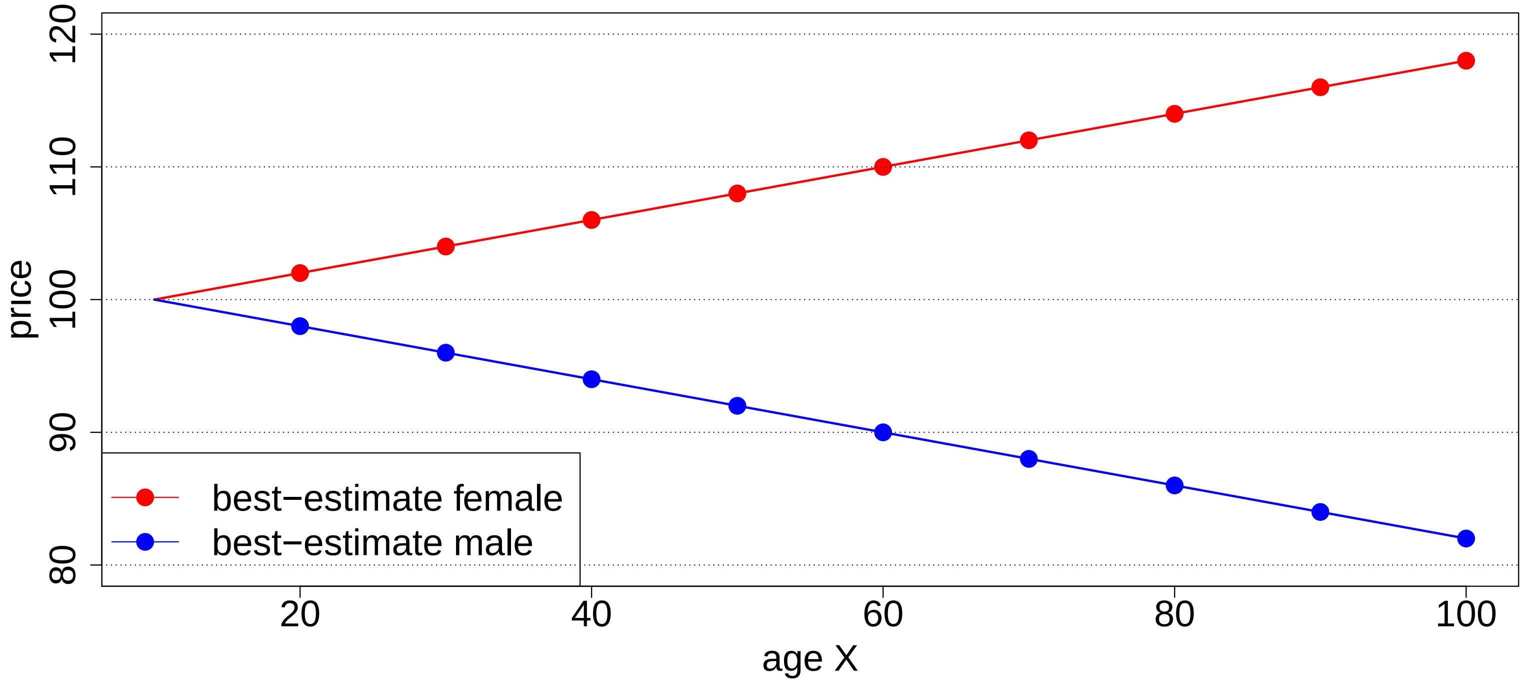
- Develop regression model for Y using covariates $\boldsymbol{X}$ and $\boldsymbol{D}$ as explanatory variables.

- This motivates best-estimate price for Y

$$\mu(\boldsymbol{X}, \boldsymbol{D}) = \mathbb{E}\left[Y \middle| \boldsymbol{X}, \boldsymbol{D}\right].$$

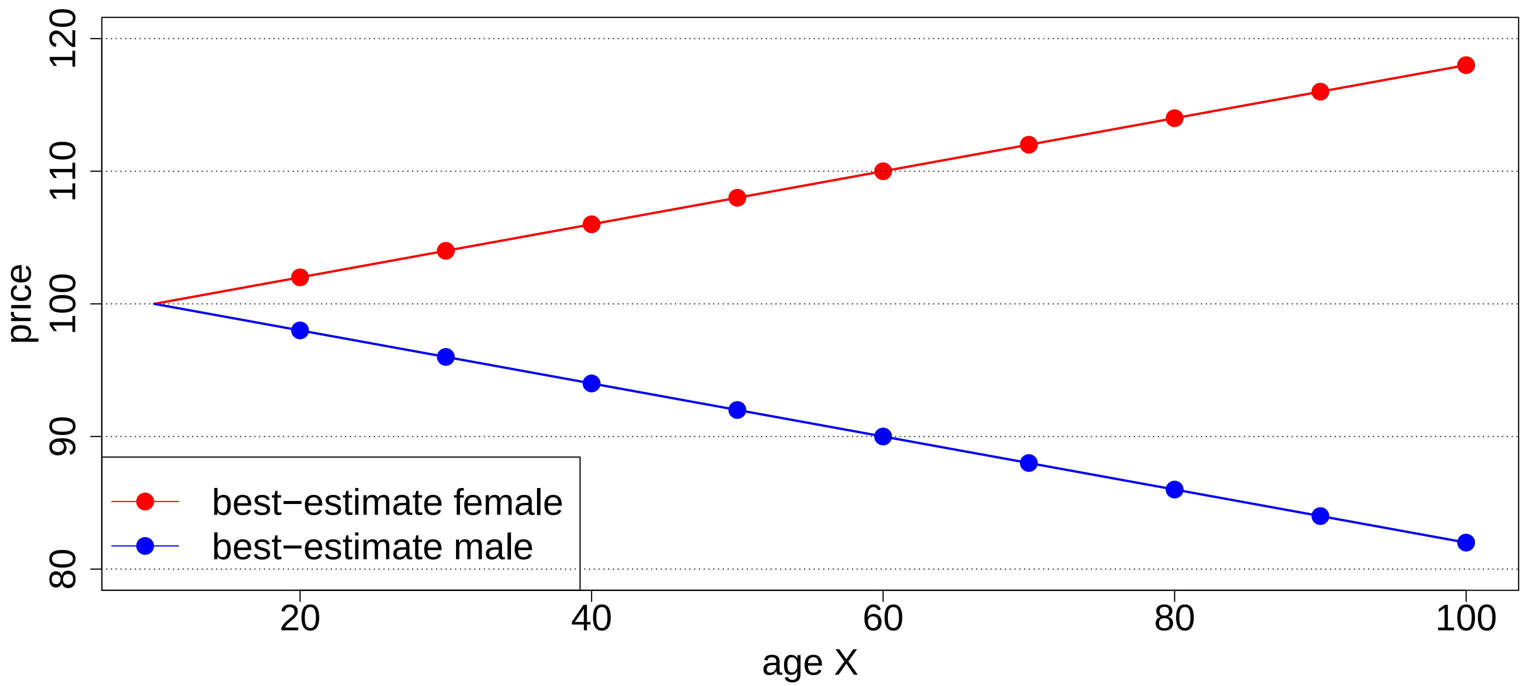
- The best-estimate price
  - ⋆ uses maximal available information $\boldsymbol{X}$ and $\boldsymbol{D}$;
  - ⋆ minimizes prediction uncertainty (in an L^2-sense);
  - ⋆ is discriminatory w.r.t. $\boldsymbol{D}$.

Best-Estimate Price: Example



- Best-estimate prices $\mu(\boldsymbol{X}, \boldsymbol{D})$ using all available information
  - ⋆ with non-discriminatory age information $\boldsymbol{X}$;
  - ⋆ with discriminatory gender information $\boldsymbol{D}$.

Best-Estimate Price: Direct Discrimination

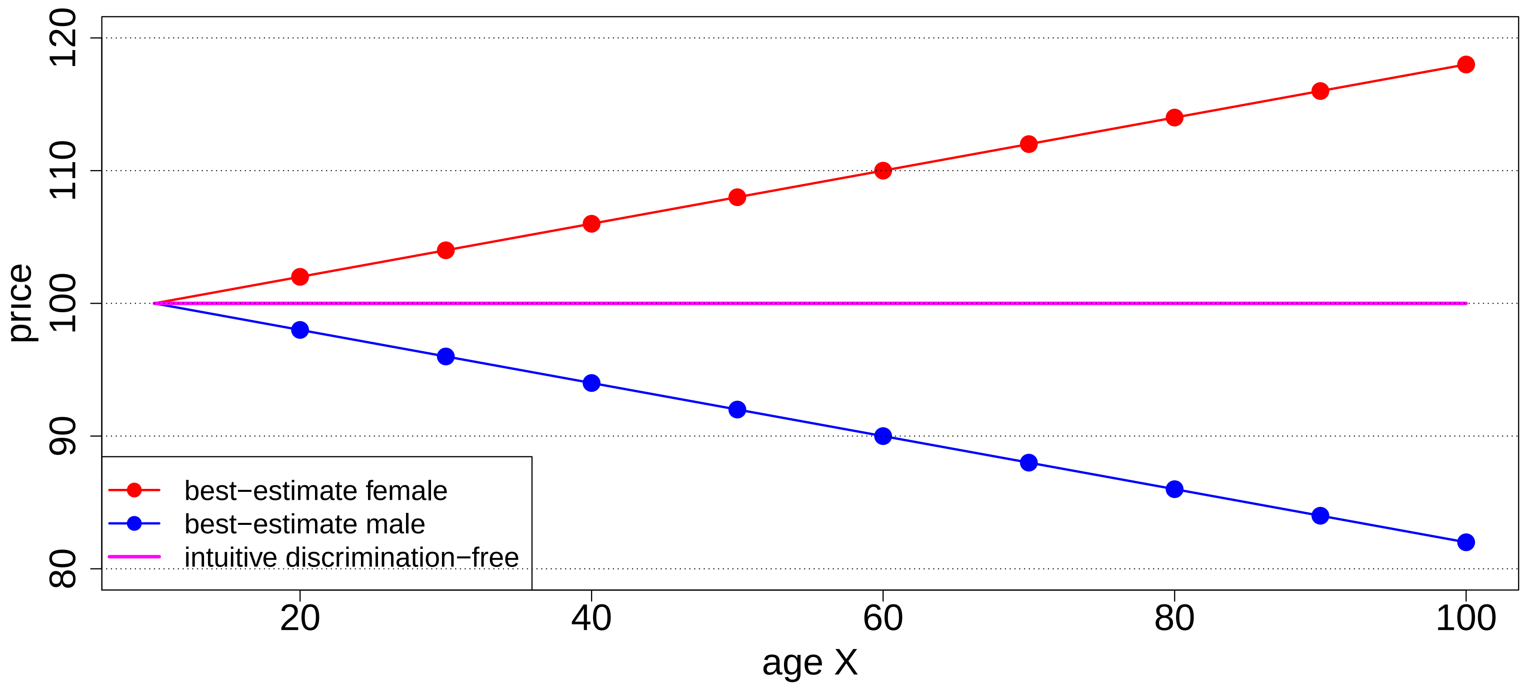


- Article 2(a):[1] "direct discrimination: where one person is treated less favourably, on grounds of sex..."

- Intuitive guess for discrimination-free price?

[1] COUNCIL DIRECTIVE 2004/113/EC of 13 December 2004, Official Journal of the European Union L 373/37

Best-Estimate Price: Direct Discrimination



- Article 2(a): "direct discrimination: where one person is treated less favourably, on grounds of sex..."

- Intuitive guess for discrimination-free price.

- **Unawareness Price and Indirect Discrimination**

Unawareness Price

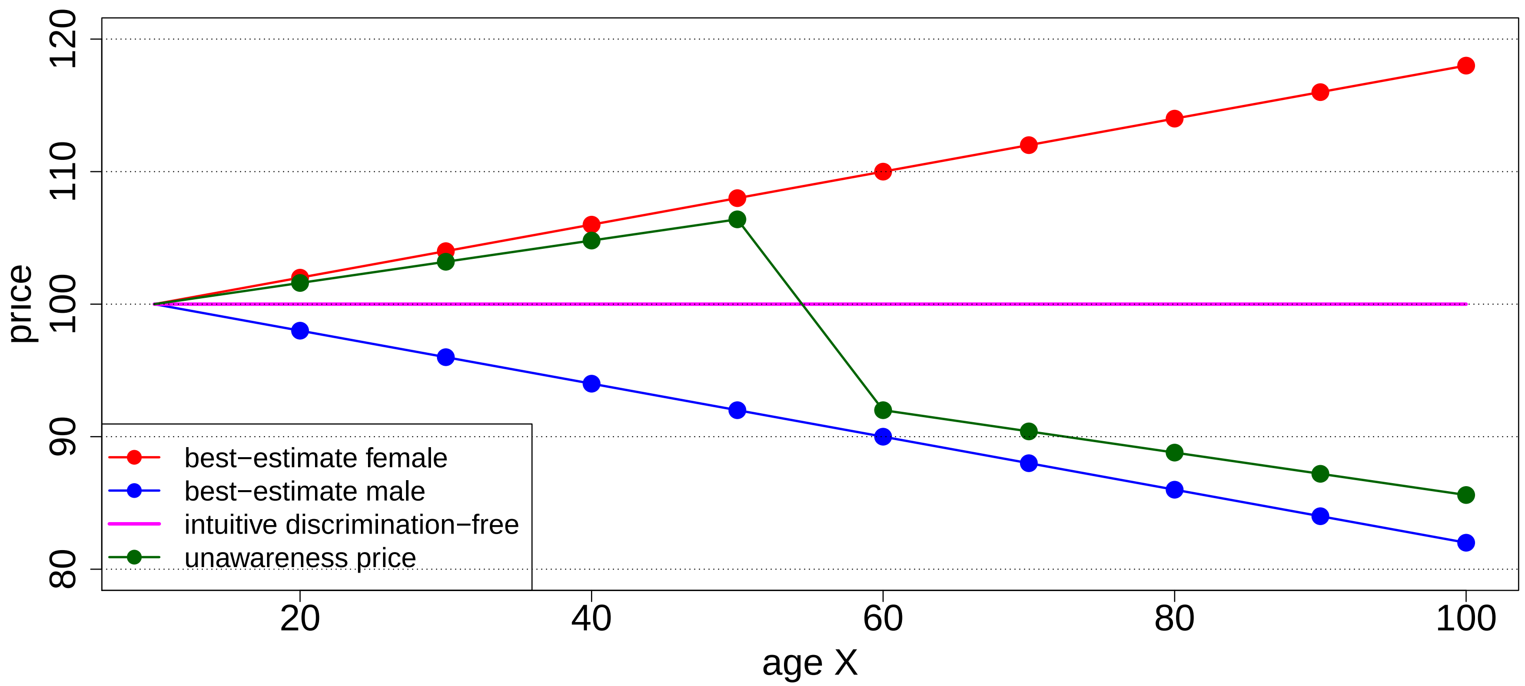
- Direct discrimination can be avoided by dropping discriminatory information $\boldsymbol{D}$.

- This provides unawareness price for Y

$$\mu(\boldsymbol{X}) = \mathbb{E}\left[Y \mid \boldsymbol{X}\right].$$

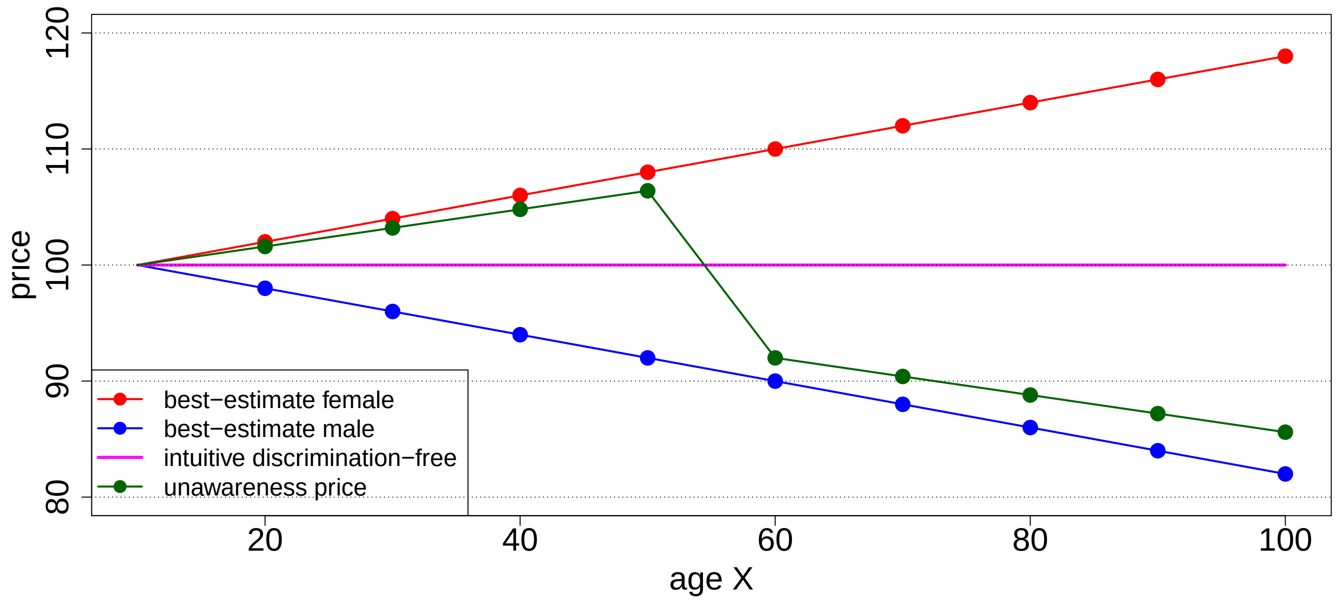
- The unawareness price
  - ⋆ uses maximal available non-discriminatory information $\boldsymbol{X}$;
  - ⋆ minimizes prediction uncertainty (in an L^2-sense w.r.t. $\boldsymbol{X}$);
  - ⋆ is the best approximation to the best-estimate price $\mu(\boldsymbol{X}, \boldsymbol{D})$;
  - ⋆ avoids direct discrimination.

Unawareness Price: Example

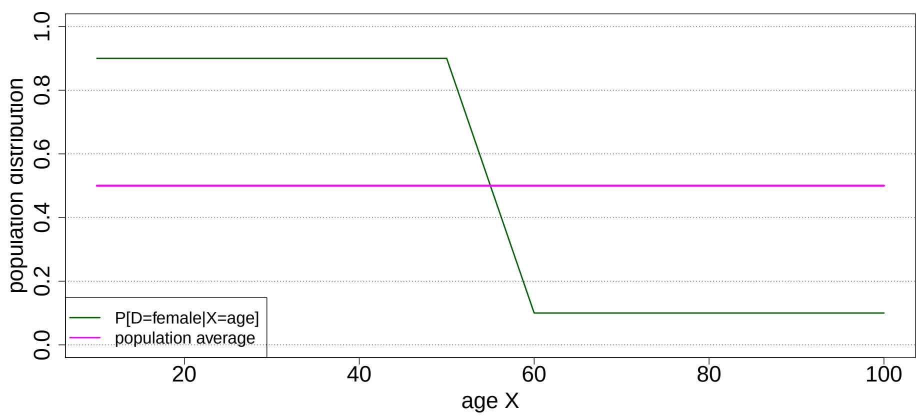


- What goes “wrong” here?

Unawareness Price: Example



- What goes “wrong” here?



What Goes "Wrong" with the Unawareness Price?

- The unawareness prices can be expressed as (tower property)

$$\begin{aligned}\mu(\boldsymbol{X}) &= \mathbb{E}\left[\mu(\boldsymbol{X},\boldsymbol{D})\right|\boldsymbol{X}] \\ &= \int_{\boldsymbol{d}} \mu(\boldsymbol{X},\boldsymbol{D}=\boldsymbol{d})\, \mathrm{d}\mathbb{P}\left(\boldsymbol{D}=\boldsymbol{d}|\boldsymbol{X}\right).\end{aligned}$$

- This shows that we infer $\boldsymbol{D}$ from $\boldsymbol{X}$ in the unawareness price.

- Article 2(b):[2] "indirect discrimination: where an apparently neutral provision... would put persons of one sex at a particular disadvantage compared with persons of the other sex, unless that provision... is objectively justified..."

[2] COUNCIL DIRECTIVE 2004/113/EC of 13 December 2004, Official Journal of the European Union L 373/37

- **Discrimination-Free Price**

Discrimination-Free Pricing

- The unawareness prices can be expressed as (tower property)

$$\begin{aligned} \mu(\boldsymbol{X}) &= \mathbb{E}\left[\mu(\boldsymbol{X},\boldsymbol{D})\right|\boldsymbol{X}] \\ &= \int_{\boldsymbol{d}} \mu(\boldsymbol{X},\boldsymbol{D}=\boldsymbol{d})\, \mathrm{d}\mathbb{P}\left(\boldsymbol{D}=\boldsymbol{d}|\boldsymbol{X}\right). \end{aligned}$$

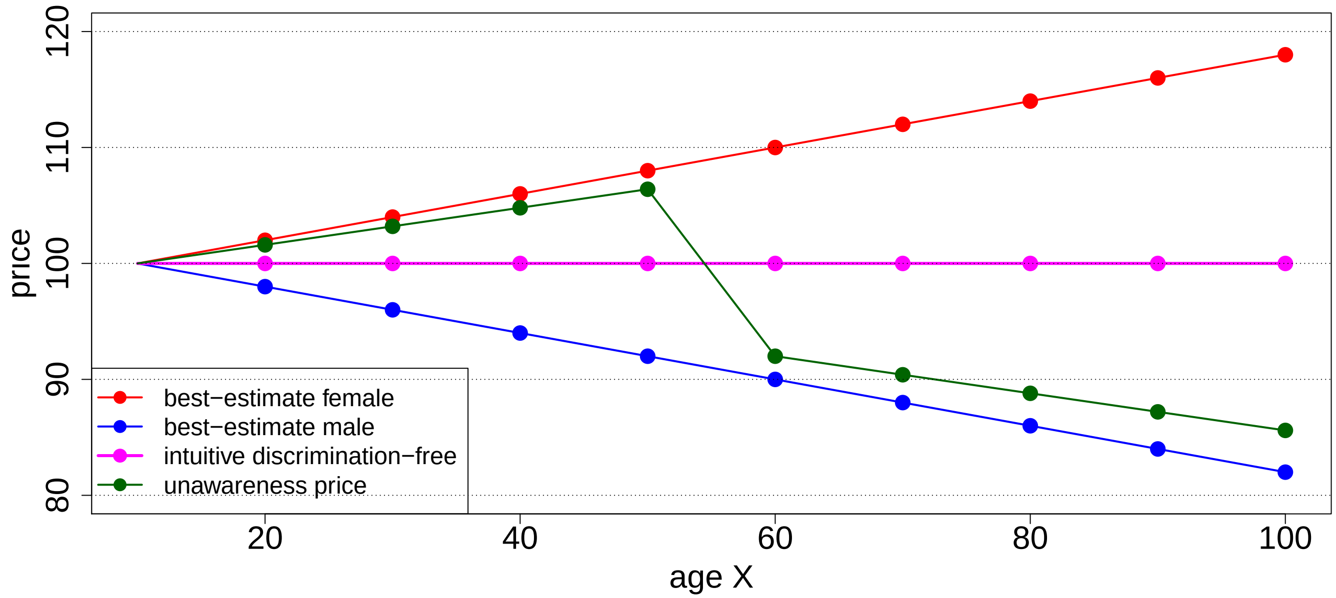
- We need to "break the structure" that allows to infer $\boldsymbol{D}$ from $\boldsymbol{X}$.

- This motivates discrimination-free price

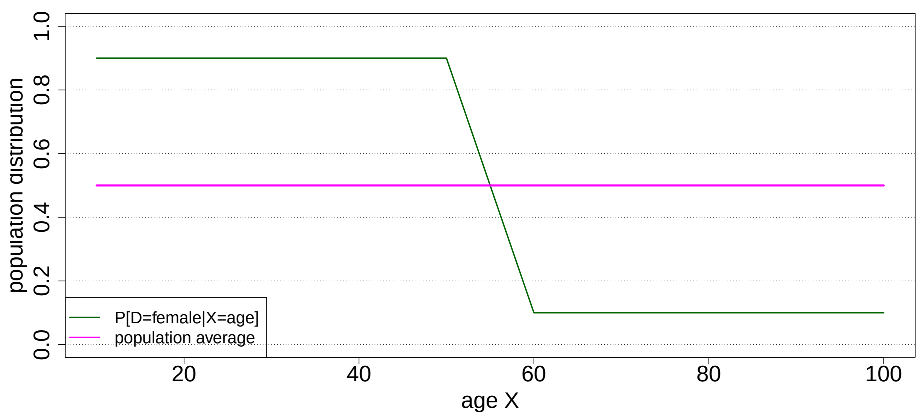
$$\mu^*(\boldsymbol{X}) = \int_{\boldsymbol{d}} \mu(\boldsymbol{X},\boldsymbol{D}=\boldsymbol{d})\, \mathrm{d}\mathbb{P}^*(\boldsymbol{D}=\boldsymbol{d}),$$

for some choice $\mathbb{P}^*$ (there are infinitely many).

Discrimination-Free Price: Example



- For population distribution $\mathbb{P}^*(\boldsymbol{D}) = \mathbb{P}(\boldsymbol{D})$.



Concluding Remarks

- We need to collect discriminatory information $\boldsymbol{D}$, otherwise we cannot calculate discrimination-free prices, i.e. just knowledge of $\boldsymbol{X}$ is not good enough.

- Lindholm et al. (2020) give a mathematical definition of (in-)direct discrimination.

- For any given problem there are infinitely many choices $\mathbb{P}^*$, and henceforth there are infinitely many discrimination-free prices.

- Discrimination-free prices need to be made unbiased.

- Discrimination-free prices sacrifice predictive power relative to unawareness prices.

- Discrimination-free prices can be motivated by "do-operators" in causal statistics (confounders), see Pearl et al. (2016).

- Discrimination-free prices have same structure as partial dependence plots (PDPs), see Zhao–Hastie (2019) and Lorentzen–Mayer (2020).

- Definition of discrimination-free prices is independent of any model.
- Discrimination-free prices may induce unwanted economic side effects like adverse selection.
- Indirect discrimination can be explained by the fact that non-discriminatory covariates are used to predict discriminatory ones. The better information we have, the more accurately this can be done.
- We did not discuss fairness nor which variables are discriminatory (ethnicity, etc.).

References

- Chen, Guillén, Vigna (2018). Solvency requirement in a unisex mortality model. ASTIN Bulletin 48/3, 1219-1243.
- Chen, Vigna (2017). A unisex stochastic mortality model to comply with EU Gender Directive. Insurance: Mathematics and Economics 73, 124-136.
- Frees, Huang (2020). The discriminating (pricing) actuary. SSRN 3592475.
- Lindholm, Richman, Tsanakas, Wüthrich (2020). Discrimination-free insurance pricing. SSRN 3520676. To appear in ASTIN Bulletin 2022.
- Lorentzen, Mayer (2020). Peeking into the black box: an actuarial case study for interpretable machine learning. SSRN 3595944.
- Pearl, Glymour, Jewell (2016). Causal Inference in Statistics: A Primer. Wiley.
- Zhao, Hastie (2019). Causal interpretations of black-box models. Journal of Business & Economic Statistics.