Discrimination-Free Pricing with Bayesian Variational Inference

Lydia J. Gabric

Joint work with Kenneth Q. Zhou and Shuang Zhou

Arizona State University

lgabric@asu.edu

Actuarial Research Conference
July 31, 2023

Agenda

Agenda

Discrimination in the late 19th century

An excerpt from Hoffman (1900, pg. 137):

NEWARK, N. J., March 10, 1881.

TO SUPERINTENDENTS AND AGENTS.

The following changes will be made with respect to colored persons (Negroes), applying for assurance in this Company, under policies issued on and after the week commencing Monday, March 28, 1881. (This applies to all APPLICATIONS taken during the week commencing Monday, March 21st.)

1. Under Adult Policies the sum assured will be ONE-THIRD less than now granted for the same weekly premium.

2. Under Infantile Policies the amount assured will be the same as now, but the weekly premium will be increased to FIVE CENTS.

These changes are made in consequence of the excessive mortality prevailing in the class above named; they do not apply to other persons. Policies issued prior to March 28th will not be affected by this regulation.

Rate tables for use with Colored Applicants will be duly sent you.

Agents using Infantile Applications in which the question of "Race" is not asked, should write on the lower margin on the back of the application the word "white" or "colored" as the case may be—unless this is done the application will be returned for correction.

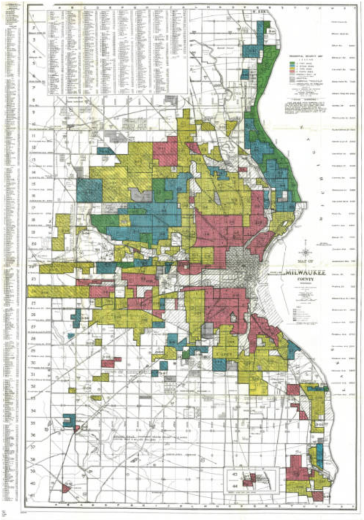
JOHN F. DRYDEN,
Secretary.

Direct discrimination as race directly affects benefits and premiums.

Discrimination at Wisconsin in 1938

- Resident Security Map of Milwaukee County, WI
- Redlining denoted areas in which banks and lenders would deny mortgage loans
- Also found within the insurance industry (Ansfield, 2021)

Indirect Discrimination from using geographic location as a proxy for race



United States. Federal Home Loan Bank Board. Division of Research and Statistics (1938)

Where is the industry now?

As of 2012, thirteen states had general statutes banning any kind of "unfair discrimination" within life, health, disability, auto, and property/casualty insurance.

- Although, these general prohibitions do not mention any specific characteristics (Avraham et al., 2014).

In 2011, the European Union Court of Justice banned gender-based discrimination in insurance.

- However, the Guidelines on the Application of the Gender Directive allowed the use of true risk factors that may be correlated with gender (European Commission, 2012).

Existing Research

- Review of discrimination in insurance:
  - Frees and Huang (2021) – Discriminating actuary
  - Dolman et al. (2021) – Multidisciplinary collaboration
- Unisex mortality models:
  - Chen and Vigna (2017) – Unisex mortality model
  - Chen et al. (2018) – Solvency requirement
- Fairness and discrimination:
  - Grari et al. (2022) – Fair pricing via adversarial learning
  - Charpentier (2022) – Quantifying fairness and discrimination
  - Lindholm et al. (2023) – Demographic disparities
  - Xin and Huang (2023) – Antidiscrimination Insurance Pricing
- Discrimination-Free Pricing:
  - Lindholm et al. (2022a) – Frequentist inference
  - Andrés et al. (2022) – Casual inference
  - Lindholm et al. (2022b) – Multi-task network

Background and Terminology

- Discrimination: “the act of treating different groups differently” (Frees & Huang, 2021)

Background and Terminology

- Discrimination: “the act of treating different groups differently” (Frees & Huang, 2021)
  - “A pricing model *avoids direct discrimination,* if none of the discriminatory features (characteristics) is used as a rating factor” (Lindholm et al., 2022a)
  - “A pricing model *avoids indirect discrimination,* if it avoids direct discrimination and, furthermore, the nondiscriminatory features are used in a way that does not allow implicit inference of discriminatory features from them” (Lindholm et al., 2022a)

Background and Terminology

- Discrimination: “the act of treating different groups differently” (Frees & Huang, 2021)
  - “A pricing model *avoids direct discrimination,* if none of the discriminatory features (characteristics) is used as a rating factor” (Lindholm et al., 2022a)
  - “A pricing model *avoids indirect discrimination,* if it avoids direct discrimination and, furthermore, the nondiscriminatory features are used in a way that does not allow implicit inference of discriminatory features from them” (Lindholm et al., 2022a)

- “Unfair discrimination exists if, after allowing for practical limitations, price differentials fail to reflect equitably the difference in expected losses and expenses” (National Association of Insurance Commissioners (NAIC), 2010)

Background and Terminology

A general definition of discrimination given by Frees and Huang (2021):

"The act of treating different groups differently"

A further categorization of the definitions of discrimination:

	Justifiable	Unjustifiable
Direct	Age	Race
Indirect	**???**	ZIP Code*

* Use ZIP Code to infer the relationship with race and use it in pricing.

Our objectives

Provide a cohesive Bayesian pricing framework that produces discrimination-free and asymptotically unbiased prices *without* the knowledge of policyholder's discriminatory information

Our objectives

Provide a cohesive Bayesian pricing framework that produces discrimination-free and asymptotically unbiased prices *without* the knowledge of policyholder's discriminatory information

- Reconstruct the discrimination-free pricing expressions of Lindholm et al. (2022a) from a Bayesian perspective
- Define Bayesian mixture models that treat the discriminatory covariates as latent variables without using individual level discriminatory information
- Implement Bayesian Variational Inference to obtain the approximate posterior distribution that achieves discrimination-free price estimates

Agenda

A Review of Lindholm et al. (2022a)

Let $\mathbf{D}$ and $\mathbf{X}$ be two vectors of the discriminatory and non-discriminatory covariates of a policyholder, respectively, and Y be the claim amount of the policyholder.

A Review of Lindholm et al. (2022a)

Let $\mathbf{D}$ and $\mathbf{X}$ be two vectors of the discriminatory and non-discriminatory covariates of a policyholder, respectively, and Y be the claim amount of the policyholder.

Best-Estimate:

$$\mu(\mathbf{X}, \mathbf{D}) := \mathbb{E}[Y|\mathbf{X}, \mathbf{D}] \tag{1}$$

A Review of Lindholm et al. (2022a)

Let $\mathbf{D}$ and $\mathbf{X}$ be two vectors of the discriminatory and non-discriminatory covariates of a policyholder, respectively, and Y be the claim amount of the policyholder.

Best-Estimate:

$$\mu(\mathbf{X}, \mathbf{D}) := \mathbb{E}[Y|\mathbf{X}, \mathbf{D}] \tag{1}$$

Unawareness:

$$\mu(\mathbf{X}) := \mathbb{E}[Y|\mathbf{X}] = \int_{\mathbf{d}} \mu(\mathbf{X}, \mathbf{d}) \mathrm{d}\mathbb{P}(\mathbf{d}|\mathbf{X}) \tag{2}$$

A Review of Lindholm et al. (2022a)

Let **D** and **X** be two vectors of the discriminatory and non-discriminatory covariates of a policyholder, respectively, and Y be the claim amount of the policyholder.

Best-Estimate:

$$\mu(\mathbf{X}, \mathbf{D}) := \mathbb{E}[Y|\mathbf{X}, \mathbf{D}] \tag{1}$$

Unawareness:

$$\mu(\mathbf{X}) := \mathbb{E}[Y|\mathbf{X}] = \int_{\mathbf{d}} \mu(\mathbf{X}, \mathbf{d}) d\mathbb{P}(\mathbf{d}|\mathbf{X}) \tag{2}$$

Biased Discrimination-Free:

$$h(\mathbf{X}) := \int_{\mathbf{d}} \mu(\mathbf{X}, \mathbf{d}) d\mathbb{P}(\mathbf{d}) \tag{3}$$

A Review of Lindholm et al. (2022a)

Let **D** and **X** be two vectors of the discriminatory and non-discriminatory covariates of a policyholder, respectively, and Y be the claim amount of the policyholder.

Best-Estimate:

$$\mu(\mathbf{X}, \mathbf{D}) := \mathbb{E}[Y|\mathbf{X}, \mathbf{D}] \tag{1}$$

Unawareness:

$$\mu(\mathbf{X}) := \mathbb{E}[Y|\mathbf{X}] = \int_{\mathbf{d}} \mu(\mathbf{X}, \mathbf{d}) d\mathbb{P}(\mathbf{d}|\mathbf{X}) \tag{2}$$

Biased Discrimination-Free:

$$h(\mathbf{X}) := \int_{\mathbf{d}} \mu(\mathbf{X}, \mathbf{d}) d\mathbb{P}(\mathbf{d}) \tag{3}$$

Unbiased Discrimination-Free:

$$h^*(\mathbf{X}) := \int_{\mathbf{d}} \mu(\mathbf{X}, \mathbf{d}) d\mathbb{P}^*(\mathbf{d}) \tag{4}$$

Agenda

Bayesian pricing: Directly using observed $\mathbf{D}_i$

Let $\boldsymbol{\Theta}$ be a vector of the unknown model parameters, and suppose that

$$Y_i \mid \mathbf{X}_i, \mathbf{D}_i, \boldsymbol{\Theta} \sim f(y_i; \mathbf{x}_i, \mathbf{d}_i, \theta), \quad i = 1, \ldots, n, \tag{5}$$
$$\boldsymbol{\Theta} \sim \pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta}),$$

where

- $f(y_i; \mathbf{x}_i, \mathbf{d}_i, \boldsymbol{\theta})$ denotes the likelihood function of Y_i for given values of $\mathbf{x}_i, \mathbf{d}_i, \boldsymbol{\theta}$, and
- $\pi(\boldsymbol{\theta})$ is the prior distribution of the unknown model parameters $\boldsymbol{\Theta}$.

Goal: Obtain the posterior distribution of $\boldsymbol{\Theta}$, $\pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|Y, \mathbf{X}, \mathbf{D})$, for Bayesian inference.

Bayesian pricing: Directly using observed $\mathbf{D}_i$

The Bayesian best-estimate is

$$\mu_B(\mathbf{X}_i, \mathbf{D}_i) = \mathbb{E}[Y_i|\mathbf{X}_i, \mathbf{D}_i] = \int y_i \cdot g(y_i|\mathbf{X}_i, \mathbf{D}_i) \cdot dy_i \tag{6}$$

where

$$g(y_i|\mathbf{X}_i, \mathbf{D}_i) = \int_{\boldsymbol{\theta}} f(y_i|\mathbf{X}_i, \mathbf{D}_i, \boldsymbol{\theta}) \cdot \pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta}$$

Bayesian pricing: Directly using observed $\mathbf{D}_i$

The Bayesian best-estimate is

$$\mu_B(\mathbf{X}_i, \mathbf{D}_i) = \mathbb{E}[Y_i|\mathbf{X}_i, \mathbf{D}_i] = \int y_i \cdot g(y_i|\mathbf{X}_i, \mathbf{D}_i) \cdot dy_i \tag{6}$$

where

$$g(y_i|\mathbf{X}_i, \mathbf{D}_i) = \int_{\boldsymbol{\theta}} f(y_i|\mathbf{X}_i, \mathbf{D}_i, \boldsymbol{\theta}) \cdot \pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta}$$

Inference Methods	Frequentist	Bayesian
Best-Estimate	$\mu(\mathbf{X}_i, \mathbf{D}_i)$	$\mu_B(\mathbf{X}_i, \mathbf{D}_i)$
Unawareness	$\mu(\mathbf{X}_i)$	$\mu_B(\mathbf{X}_i)$
Discrimination-Free	$h(\mathbf{X}_i)$	$h_B(\mathbf{X}_i)$
Unbiased Discrimination-Free	$h^*(\mathbf{X}_i)$	$h_B^*(\mathbf{X}_i)$

Table: Reconstruction of Lindholm et al. (2022a) under Bayesian inference

Bayesian pricing: Indirectly using latent $\mathbf{D}_i$

Assume that the discriminatory covariates $\mathbf{D}_i$ are **not observed**. Let $\boldsymbol{\Theta}$ be a vector of the unknown model parameters, and suppose

$$Y_i \mid \mathbf{X}_i, \mathbf{D}_i, \boldsymbol{\Theta} \sim f(y_i; \mathbf{x}_i, \mathbf{d}_i, \theta),$$

$$\mathbf{D}_i \overset{\text{iid}}{\sim} \pi_{\mathbf{D}_i}(\mathbf{d}_i), \quad \boldsymbol{\Theta} \sim \pi_{\boldsymbol{\Theta}}(\theta), \quad i = 1, \ldots, n,$$

where

- $f(y_i; \mathbf{x}_i, \mathbf{d}_i, \theta)$ denotes the likelihood function of Y_i,
- $\pi_{\mathbf{D}_i}(\mathbf{d}_i)$ is the prior distribution for the latent discriminatory variables, and
- $\pi_{\boldsymbol{\Theta}}(\theta)$ is the prior distribution for the unknown model parameters.

Goal: Obtain the posterior distribution of $\mathbf{D}$ and $\boldsymbol{\Theta}$, $\pi(\mathbf{d}, \theta | Y, \mathbf{X})$, for Bayesian inference.

Bayesian Inference

To obtain the posterior distribution of $\mathbf{D}$ and $\mathbf{\Theta}$:

Bayesian Inference

To obtain the posterior distribution of $\mathbf{D}$ and $\boldsymbol{\Theta}$:

a. Markov chain Monte Carlo (MCMC): sequential **sampling** established by a selected simulation algorithm (Gelman et al., 2013).
   - When a parameter is updated, it is conditioned on the most recent value of the rest of the parameters.
   - While conditional sampling $\mathbf{D}|Y, \mathbf{X}, \boldsymbol{\Theta}$, we are able to infer $\mathbf{D}$ with the observed data $\{Y, \mathbf{X}\}$ and the most recent values of $\boldsymbol{\Theta}$ which results in indirect discrimination.

Bayesian Inference

To obtain the posterior distribution of $\mathbf{D}$ and $\boldsymbol{\Theta}$:

a. Markov chain Monte Carlo (MCMC): sequential **sampling** established by a selected simulation algorithm (Gelman et al., 2013).
   - When a parameter is updated, it is conditioned on the most recent value of the rest of the parameters.
   - While conditional sampling $\mathbf{D}|Y, \mathbf{X}, \boldsymbol{\Theta}$, we are able to infer $\mathbf{D}$ with the observed data $\{Y, \mathbf{X}\}$ and the most recent values of $\boldsymbol{\Theta}$ which results in indirect discrimination.

b. Variational Inference (VI): approximate posterior distribution through **optimization** (Blei et al., 2017).
   - Consider a family of distributions, $\mathcal{Q}$, of the unknown variables, $\mathbf{z} = \{\mathbf{D}, \boldsymbol{\Theta}\}$, and select the member from the family that minimizes the KL divergence to the true posterior distribution:

$$q^*(\mathbf{z}) = \underset{q \in \mathcal{Q}}{\operatorname{argmin}} \ \mathrm{KL}(q(\mathbf{z}) || p(\mathbf{z}|\mathbf{w})),$$

where $\mathbf{w} = \{Y, \mathbf{X}\}$ contains all of the observed variables.

Bayesian Price Estimates

	Justifiable (Discrimination-Free)	Unjustifiable (Discriminatory)
Direct (Observed **D**)	$h_B(\mathbf{X})$ and $h_B^*(\mathbf{X})$	$\mu_B(\mathbf{X}, \mathbf{D})$ and $\mu_B(\mathbf{X})$
Indirect (Latent **D**)	Variational Inference	MCMC Sampling

Remarks:

- The unawareness, biased, and unbiased discrimination-free price are all functions of the best estimate prices.
- The MCMC method implicitly infers the discriminatory information D through conditional sampling.
- Variational inference with mean-field approximation assumes independence between all unknown variables.

Agenda

Data and Data Simulation

Considered the swmotorcycle dataset from the CASdatasets $\mathbb{R}$ package (Dutang & Charpentier, 2020). We selected the following fields:

- Y: claim count
- X: risk class $= \{Low, Medium, High\}$
- D: gender $= \{Male, Female\}$

Data and Data Simulation

Considered the swmotorcycle dataset from the CASdatasets R package (Dutang & Charpentier, 2020). We selected the following fields:

- Y: claim count
- X: risk class $= \{Low, Medium, High\}$
- D: gender $= \{Male, Female\}$

Data simulation procedure:

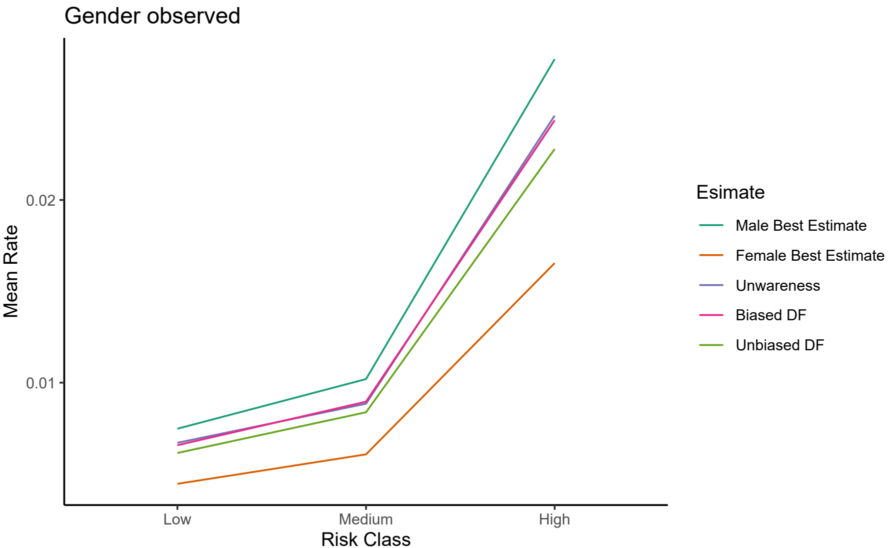
1. General cleaning and obtained 20,000 observations
2. Fit a Poisson glm with log link to obtain true model coefficients

$$\ln Y = \beta_0 + \beta_1 \mathbb{1}_{\{X=Medium\}} + \beta_2 \mathbb{1}_{\{X=High\}} + \beta_3 \mathbb{1}_{\{D=Female\}}$$

3. Simulated a claim count realization for each observation with the true model coefficients

Direct use of gender

Bayesian reconstruction of framework established by Lindholm et al. (2022a) using observed gender.



Indirect use of gender

Bayesian model that assumes gender is a latent variable.

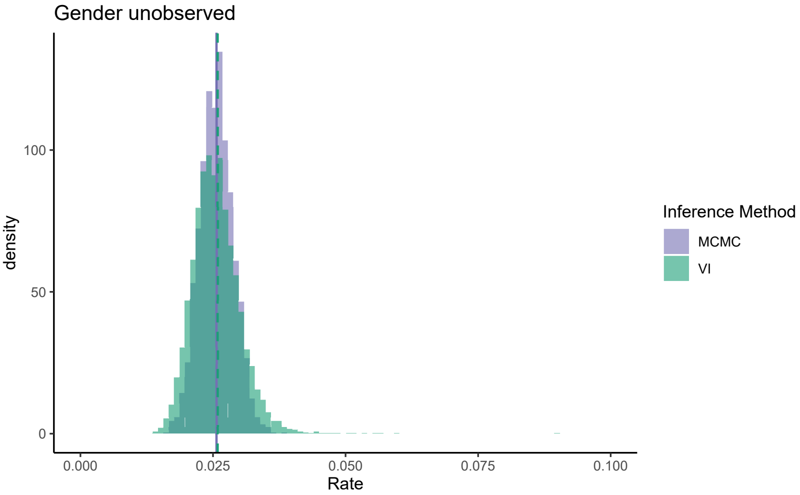


Figure: Bayesian estimates of the claim rate for the high risk class

Agenda

Overview

We proposed a Bayesian pricing model that does not require the use of individual-level discriminatory data and provides a direct estimation process to obtain asymptotically unbiased discrimination-free prices.

- We reconstructed the discrimination-free pricing framework proposed by Lindholm et al. (2022a) from a Bayesian perspective
- Through the use of Bayesian mixture models and latent variables, our model does not require discriminatory information on the individual level.
- Using Bayesian variational inference with a mean-field approximation, we ensure discrimination-free via the assumed independence between discriminatory variables and model parameters.

References I

Andrés, C., Iturria, A., Hardy, M., & Marriott, P. (2022). *A discrimination-free premium under a causal framework*. https://ssrn.com/abstract=4079068

Ansfield, B. (2021). The Crisis of Insurance and the Insuring of the Crisis: Riot Reinsurance and Redlining in the Aftermath of the 1960s Uprisings. *Journal of American History*, *107*(4), 899–921. https://doi.org/10.1093/jahist/jaaa533

Avraham, R., Logue, K. D., & Schwarcz, D. (2014). Understanding insurance antidiscrimination laws. *Southern California law review*, *87*(2), 195–274.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Charpentier, A. (2022). Quantifying fairness and discrimination in predictive models. http://arxiv.org/abs/2212.09868

Chen, A., Guillen, M., & Vigna, E. (2018). Solvency requirement in a unisex mortality model. *ASTIN Bulletin*, *48*, 1219–1243. https://doi.org/10.1017/asb.2018.11

Chen, A., & Vigna, E. (2017). A unisex stochastic mortality model to comply with eu gender directive. *Insurance: Mathematics and Economics*, *73*, 124–136. https://doi.org/10.1016/j.insmatheco.2017.01.007

References II

Dolman, C., Frees, E., & Huang, F. (2021). *Multidisciplinary collaboration on discrimination - not just nice to have*. https://doi.org/10.1017/S174849952100021X

Dutang, C., & Charpentier, A. (2020). *Casdatasets: Insurance datasets* [R package version 1.0-11].

European Commission. (2012). Guidelines on the application of council directive 2004/113/ec to insurance, in the light of the judgment of the court of justice of the european union in case c-236/09 (test-achats). http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1416170084502&uri=CELEX:32014R0269

Frees, E. W., & Huang, F. (2021). The discriminating (pricing) actuary. *North American Actuarial Journal*. https://doi.org/10.1080/10920277.2021.1951296

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis, third edition, 3rd edition* (3rd edition). CRC Press.

Grari, V., Charpentier, A., & Detyniecki, M. (2022). A fair pricing model via adversarial learning. http://arxiv.org/abs/2202.12008

References III

Hoffman, F. L. (1900). *History of the prudential insurance company of america (industrial insurance) 1875-1900*. Prudential Press.

Lindholm, M., Richman, R., Tsanakas, A., & Wüthrich, M. V. (2022a). Discrimination-free insurance pricing. *ASTIN Bulletin*, *52*, 55–89. https://doi.org/10.1017/asb.2021.23

Lindholm, M., Richman, R., Tsanakas, A., & Wüthrich, M. V. (2022b). A multi-task network approach for calculating discrimination-free insurance prices.

Lindholm, M., Richman, R., Tsanakas, A., & Wüthrich, M. V. (2023). What is fair? proxy discrimination vs. demographic disparities in insurance pricing. https://ssrn.com/abstract=4436409

National Association of Insurance Commissioners (NAIC). (2010). Property and casualty model rating law (file and use version).. https://content.naic.org/sites/default/files/GL1775.pdf

References IV

United States. Federal Home Loan Bank Board. Division of Research and Statistics. (1938). Map of milwaukee county, wisconsin : Residential security map / compiled from official records and drawn by maurice kranyecz ; published and sold by david white co., inc. ; prepared by division of research and statistics with the co-operation of the appraisal department home owners' loan corporation.. *Home Owners' Loan Corporation. Appraisal Department; Davide Whte Co., Inc.*

Xin, X., & Huang, F. (2023). Antidiscrimination insurance pricing: Regulations, fairness criteria, and models. *North American Actuarial Journal*, *0*(0), 1–35. https://doi.org/10.1080/10920277.2023.2190528

Thank You!

Questions? Comments?