

Équité en assurance : Proposition de solution en corrigéant la fonction de perte

Équipe du CDA

Mathieu Bazinet, Simon-Olivier Lépine, Liliane Ouedraogo, Michaël Rioux

23 novembre 2022



Qui sommes-nous ?



Mathieu, maîtrise en informatique avec mémoire



Liliane, baccalauréat en statistique



Simon-Olivier, maîtrise en statistique avec mémoire



Michaël, maîtrise en informatique - intelligence artificielle

Plan de la présentation

1. Problématique
2. Statistiques descriptives
3. Démarche proposée
4. Expériences
5. Conclusion

Problématique

Discrimination en assurance

- Les attributs influencent le montant de réclamation qu'on prédit.

Discrimination en assurance

- Les attributs influencent le montant de réclamation qu'on prédit.
- On ne veut pas que certains attribut protégés influencent la prime.

Discrimination en assurance

- Les attributs influencent le montant de réclamation qu'on prédit.
- On ne veut pas que certains attribut protégés influencent la prime.
- En général, il existe une corrélation entre la variable protégée et les autres variables, ce qui mène à de la discrimination indirecte. **On ne peut pas simplement retirer l'attribut protégé.**

Discrimination en assurance

- Les attributs influencent le montant de réclamation qu'on prédit.
- On ne veut pas que certains attribut protégés influencent la prime.
- En général, il existe une corrélation entre la variable protégée et les autres variables, ce qui mène à de la discrimination indirecte. **On ne peut pas simplement retirer l'attribut protégé.**
- L'estimateur d'un modèle prédictif peut avoir un biais (statistique) et un décalage (biais éthique). On veut diminuer le décalage.

Discrimination en assurance

- Les attributs influencent le montant de réclamation qu'on prédit.
- On ne veut pas que certains attribut protégés influencent la prime.
- En général, il existe une corrélation entre la variable protégée et les autres variables, ce qui mène à de la discrimination indirecte. **On ne peut pas simplement retirer l'attribut protégé.**
- L'estimateur d'un modèle prédictif peut avoir un biais (statistique) et un décalage (biais éthique). On veut diminuer le décalage.
- Diminuer le niveau de discrimination du modèle peut diminuer ses performances prédictives.

Contexte (Actulab)

- Définition mathématique de l'équité (possiblement qu'il peut être intéressant d'en comparer plus d'une).

Contexte (Actulab)

- Définition mathématique de l'équité (possiblement qu'il peut être intéressant d'en comparer plus d'une).
- Sur quels attributs est-il correct de discriminer ? Sur quels attributs est-il interdit de discriminer ?

Contexte (Actulab)

- Définition mathématique de l'équité (possiblement qu'il peut être intéressant d'en comparer plus d'une).
- Sur quels attributs est-il correct de discriminer ? Sur quels attributs est-il interdit de discriminer ?
- Que pourrions nous faire si un attribut présentement acceptable, devenait prohibé ?

Contexte (Actulab)

- Définition mathématique de l'équité (possiblement qu'il peut être intéressant d'en comparer plus d'une).
- Sur quels attributs est-il correct de discriminer ? Sur quels attributs est-il interdit de discriminer ?
- Que pourrions nous faire si un attribut présentement acceptable, devenait prohibé ?
- Comment éviter la discrimination indirecte ?

Qu'est-ce que l'équité ? (intuition)



Anna Wilson

Gender : F

veh_body = SEDAN

veh_age = 1 year

Area : A

Premium : 500\$

Pour un attribut protégé (le genre par exemple), *ceteris paribus*, nous aimerions qu'Anna et Tim aient la même prime d'assurance.



Tim Robson

Gender : M

veh_body = SEDAN

veh_age = 1 year

Area : A

Premium : 700\$

Données dataCar

- Données de 2004-2005 en Australie
- Variable binaire : présence d'une réclamation (clm)
- Variable continue : montant de la réclamation (claimcst0)
- Attribut protégé : genre (homme/femme)
- Variables explicatives : valeur du véhicule, type de véhicule, âge du véhicule, zone de résidence

Statistiques descriptives

Statistiques descriptives (présence de réclamation)

- Nombre d'observations : 67856
- 57% femmes, 43% hommes

Tableau – Répartition du nombre de réclamations dans les données dataCar

Nombre de réclamations	0	1	2	3	4
Pourcentage cumulatif	93.19	99.58	99.971	99.998	100

Statistiques descriptives (présence de réclamation)

Tableau – Pourcentage de réclamations selon le sexe

	F	M
Pas réclamation (0)	93.14	93.25
Réclamation (1)	6.86	6.75

Statistiques descriptives (Montants réclamés)

Tableau – Répartition des montants réclamés

	Min	Q_1	Q_2	Moy	Q_3	Max
Incluant 0 (67856 obs.)	0	0	0	137	0	55922
Excluant 0 (4559 obs.)	200	354	762	2014	1998	55922

Statistiques descriptives (Montants réclamés)

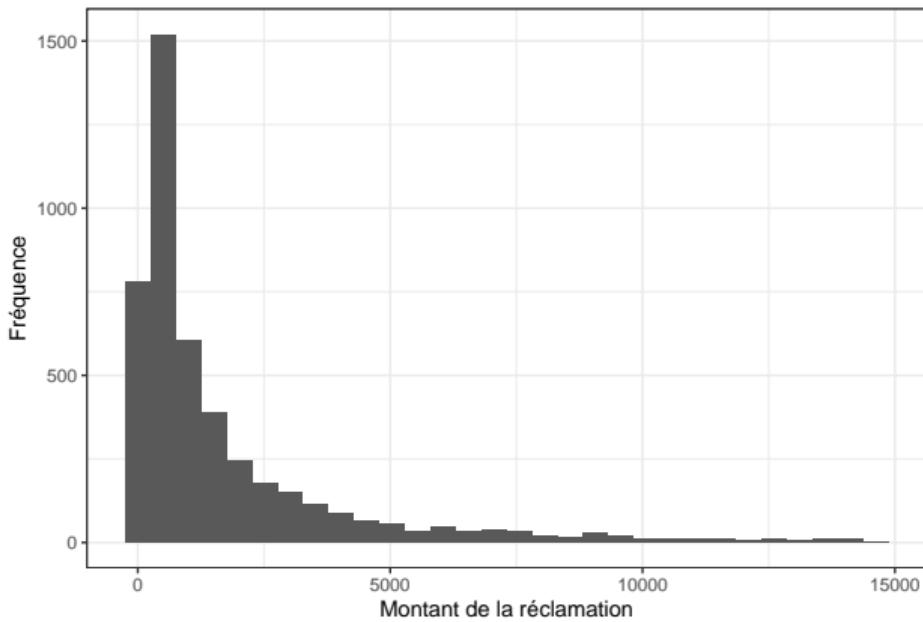
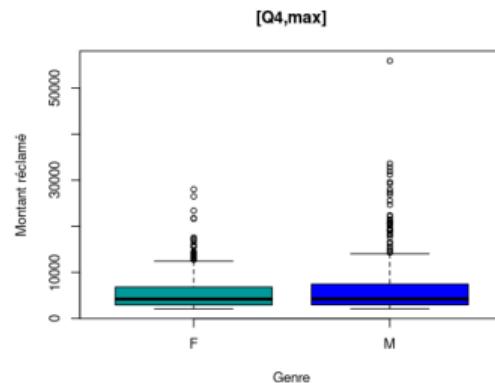
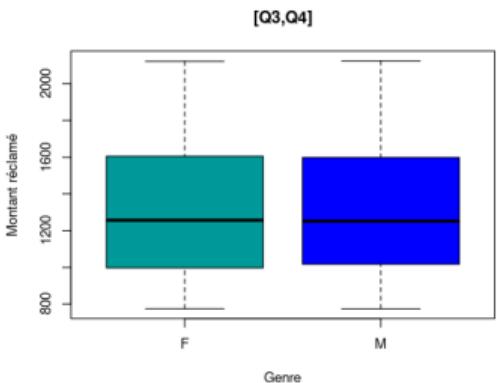
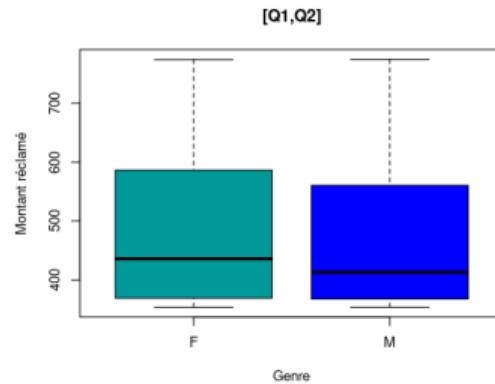
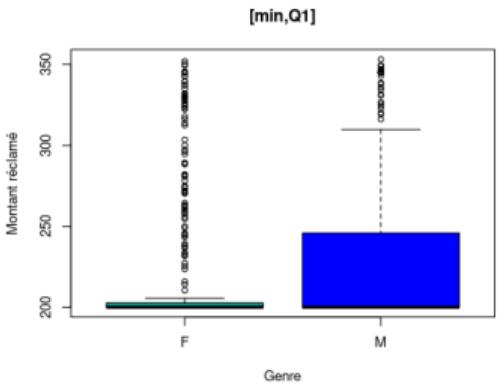


Figure – Répartition des montants réclamés, excluant 0 et les montants supérieurs à 15000\$

Statistiques descriptives (Montants réclamés)



Que faire dans le cas de la prohibition d'un attribut ?

- Dans le cas où il est complètement interdit de garder l'information en banque, on ne peut rien faire. On espère ne pas faire de discrimination indirecte.

Que faire dans le cas de la prohibition d'un attribut ?

- Dans le cas où il est complètement interdit de garder l'information en banque, on ne peut rien faire. On espère ne pas faire de discrimination indirecte.
- Notre approche permet d'offrir une solution au cas où il est possible de stocker l'information pour éviter la discrimination !

Démarche proposée

Approches possibles

- *Pre-processing* : Modification des données avant l'entraînement
- *Post-processing* : Modification du modèle après l'entraînement
- Modification du modèle durant l'entraînement

Approches possibles

- *Pre-processing* : Modification des données avant l'entraînement
- *Post-processing* : Modification du modèle après l'entraînement
- Modification du modèle durant l'entraînement

Note : Variable protégée catégorielle, soit $A \in \{a, b\}$

Modèle prédictif

- Utilisation de modèles linéaires généralisés (GLM)
 - Interprétables et utilisés par les actuaires

Modification de la fonction de perte

Fonction objectif à minimiser

$$\min_{\beta} -\ell(\beta, \phi; \mathbf{y}, \mathbf{x}) + \lambda \mathcal{D}(\beta, \phi; \mathbf{y}, \mathbf{x}), \quad (1)$$

où λ est un paramètre de pénalité qui détermine l'importance entre l'adéquation aux données (vraisemblance) et l'équité ($\mathcal{D}(\cdot)$).

Détails techniques

- Implémentation d'une régression logistique, une régression poisson et une régression gamma

Détails techniques

- Implémentation d'une régression logistique, une régression poisson et une régression gamma
- Minimisation de la fonction objectif (1) avec `scipy.optimize.minimize(..., method="BGFS")`

Détails techniques

- Implémentation d'une régression logistique, une régression poisson et une régression gamma
- Minimisation de la fonction objectif (1) avec `scipy.optimize.minimize(..., method="BGFS")`
- Choix du paramètre de régularisation λ sur les données de validation

Détails techniques

- Implémentation d'une régression logistique, une régression poisson et une régression gamma
- Minimisation de la fonction objectif (1) avec `scipy.optimize.minimize(..., method="BGFS")`
- Choix du paramètre de régularisation λ sur les données de validation
- GLM Gamma : paramètre α (*shape*) déterminé à priori sur un modèle sans régularisation pour simplifier. $\hat{\alpha}^{-1} = \chi^2/(n - p)$ (méthode des résidus de Pearson, [2])

Disparités moindres (Equalized Odds)

Disparité moindres

$$\begin{cases} \text{TVP}_{\text{hommes}} = \text{TVP}_{\text{femmes}} \\ \text{TFP}_{\text{hommes}} = \text{TFP}_{\text{femmes}} \end{cases} \quad (2)$$

Bon pour une variable discrète, mais que faire dans le cas continu, dans un contexte d'assurance ?

Parité actuarielle par groupe

Proposée par Dolman et Semenovich [1] puisque d'autres métriques n'ont aucun sens en actuariat.

PAG

$$\left| \mathbb{E} \left[\hat{M} \mid K(n-1) \leq R < Kn, A = a \right] - \mathbb{E} \left[\hat{M} \mid K(n-1) \leq R < Kn, A = b \right] \right| < \delta \quad (3)$$

pour tout groupe de taille K qui partitionne le risque des clients, \hat{M} le montant prédict de la réclamation et R le vrai risque (choisi comme étant le montant réclamé).

Parité **Actulab** par quartile

PAQ

$$\left| \mathbb{E} \left[\hat{M} \mid Q_\alpha \leq R < Q_{\alpha+\frac{1}{4}}, A = a \right] - \mathbb{E} \left[\hat{M} \mid Q_\alpha \leq R < Q_{\alpha+\frac{1}{4}}, A = b \right] \right| < \delta (Q_{\alpha+\frac{1}{4}} - Q_\alpha),$$

où Q_α sont les quartiles des montants réclamés, R est le montant réclamé, A est l'attribut protégé (binaire), \hat{M} est le montant réclamé prédict et $\delta \geq 0$.

Intuition pour le cas continu

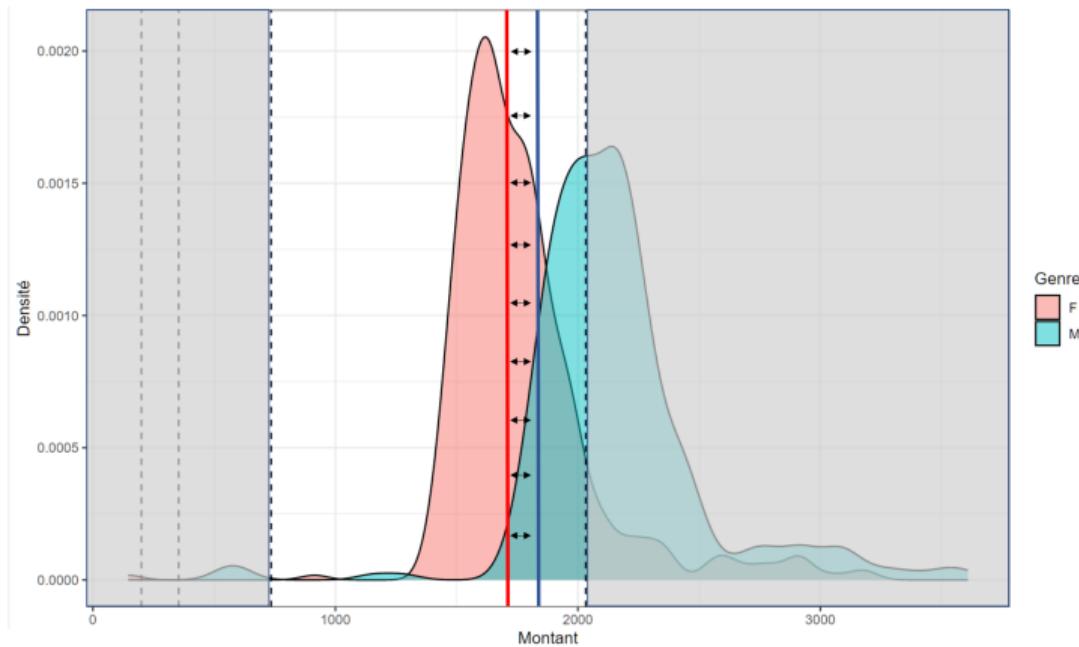


Figure – Écart entre les espérances des montant réclamés pour les hommes et les femmes entre deux quartiles

Pénalisations utilisées

Equalized Odds (variable réponse binaire)

$$\mathcal{D}_{\text{EO}} = \sum_{y \in \{+,-\}} \sum_{\substack{a,b \in A \\ a \neq b}} \left| \mathbb{P} \left[\hat{Y} = y \mid Y = y, A = a \right] - \mathbb{P} \left[\hat{Y} = y \mid Y = y, A = b \right] \right|, \quad (4)$$

où Y est la présence d'une réclamation et A est l'attribut protégé (binaire).

Pénalisations utilisées

Parité actulab par quartile (PAQ) (variable réponse continue)

$$\mathcal{D}_{\text{PAQ}} = \sum_{i=0}^3 \sum_{\substack{a,b \in A \\ a \neq b}} \frac{\left| \mathbb{E} [\hat{M} | Q_i \leq R < Q_{i+1}, A = a] - \mathbb{E} [\hat{M} | Q_i \leq R < Q_{i+1}, A = b] \right|}{Q_{i+1} - Q_i}, \quad (5)$$

où R est le montant réclamé, Q_i sont les quartiles des montants réclamés, A est l'attribut protégé (binaire) et \hat{M} est le montant réclamé prédit.

Expériences

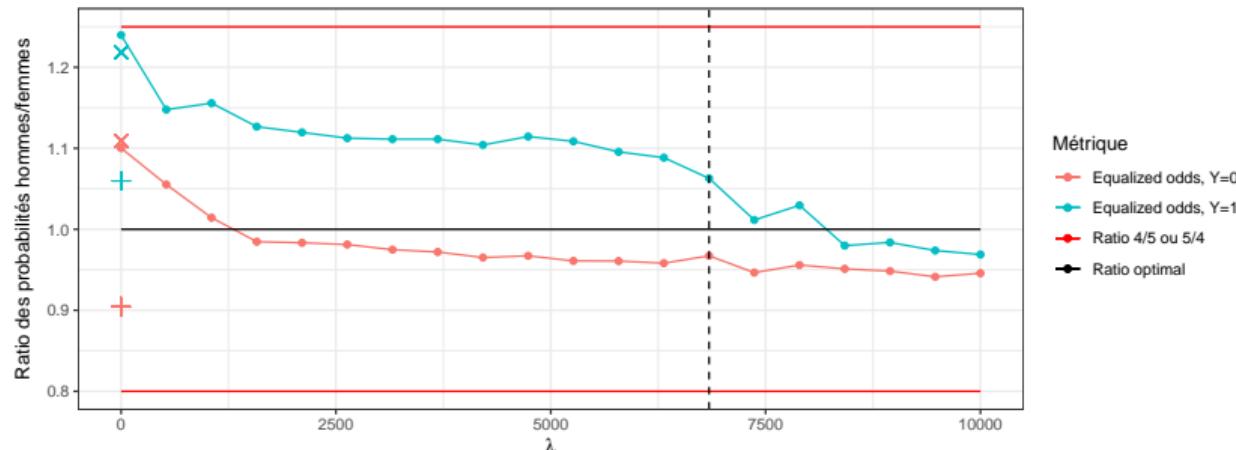
Modèles comparés

- Discrimination directe
- Discrimination indirecte (*fairness through unawareness*)
- Modèles pénalisés selon différentes valeurs de λ
 - Objectif : minimiser la discrimination

Performance prédictives des modèles

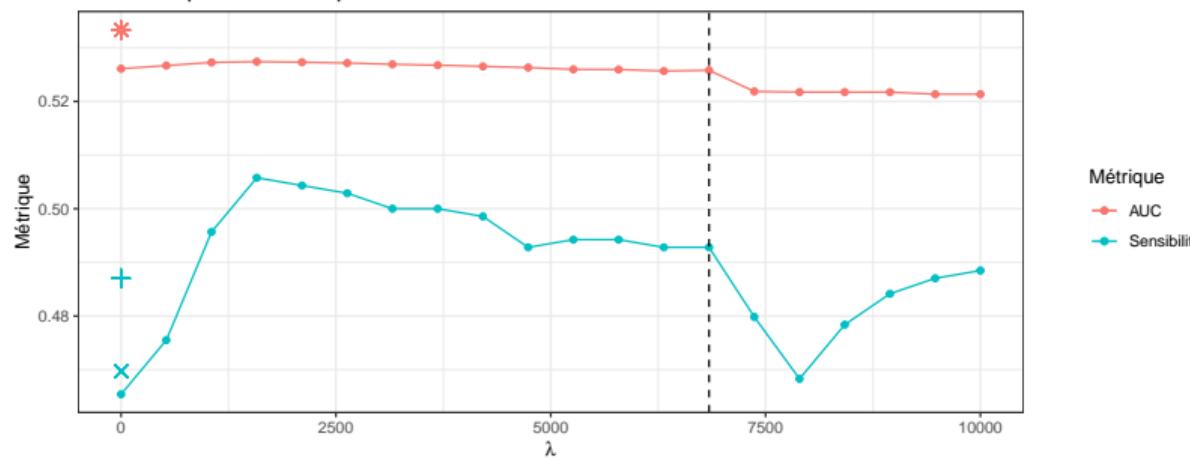
- Variable réponse binaire : AUC et sensibilité (seuil fixé)
- Variable réponse continue : NRMSE

Effet de la pénalité sur l'équité



+ : discr.
directe
× : discr.
indirecte

Effet de la pénalité sur la performance



Intuition de la métrique

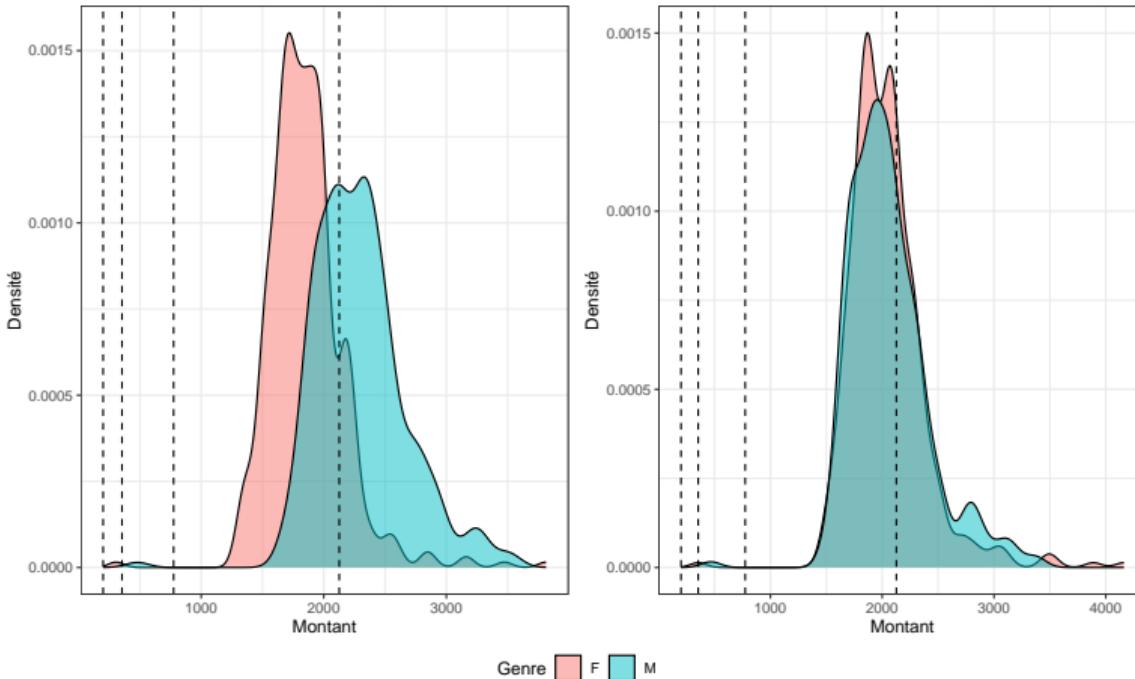


Figure – Répartition des prédictions (excluant 0) homme et femme pour la discrimination directe (gauche) et indirecte (droite)

Intuition de la métrique (suite)

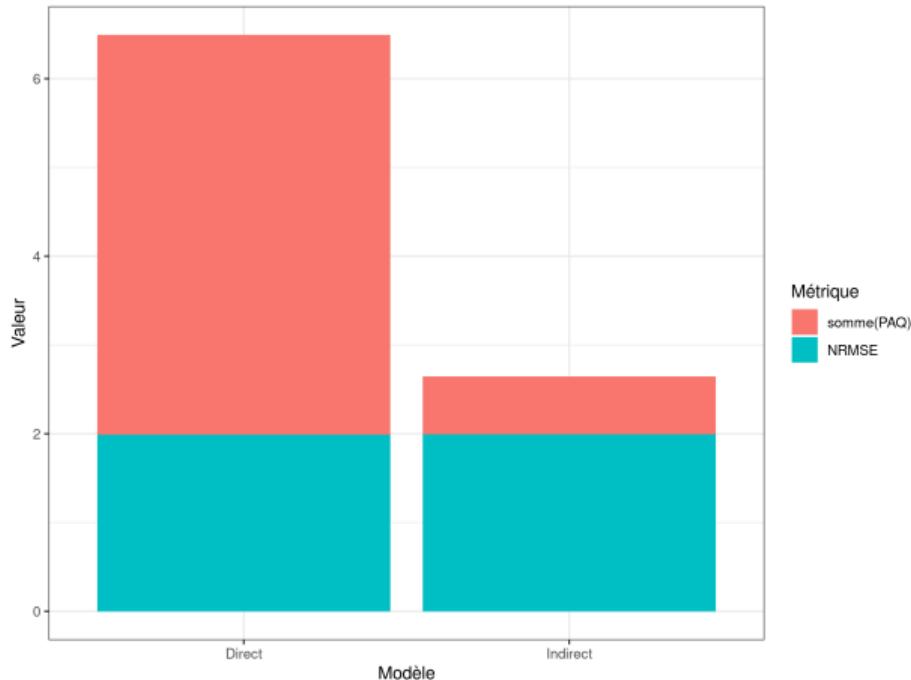


Figure – NRMSE et PAQ pour le modèle de discrimination directe et indirecte (excluant 0)

Résultats



+ : discr.
directe
x : discr.
indirecte

Résultats (suite)

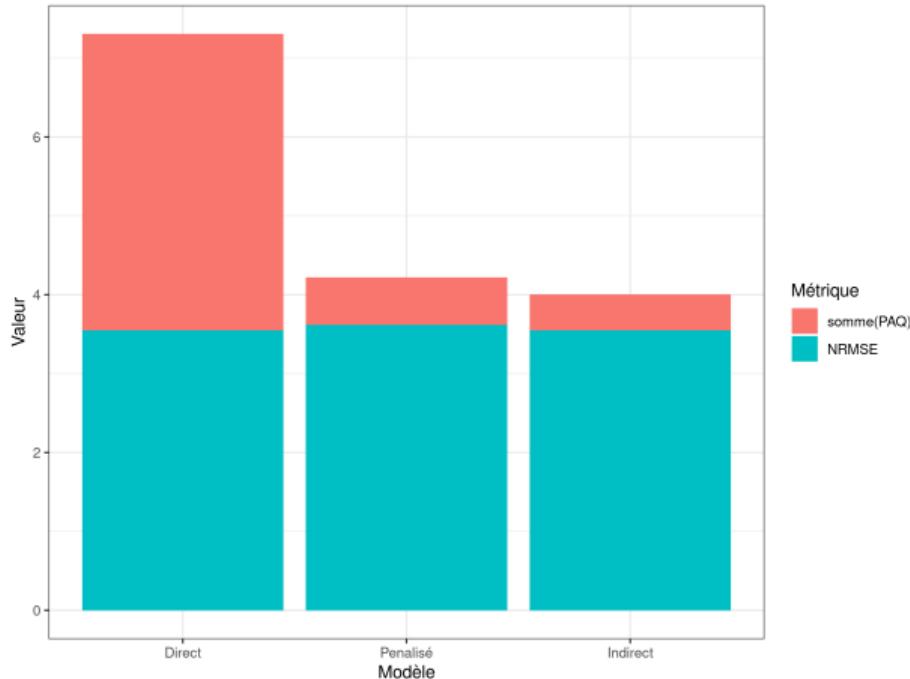


Figure – NRMSE et PAQ pour le modèle de discrimination directe, indirecte et modèle pénalisé (test)

Prédictions de valeurs extrêmes

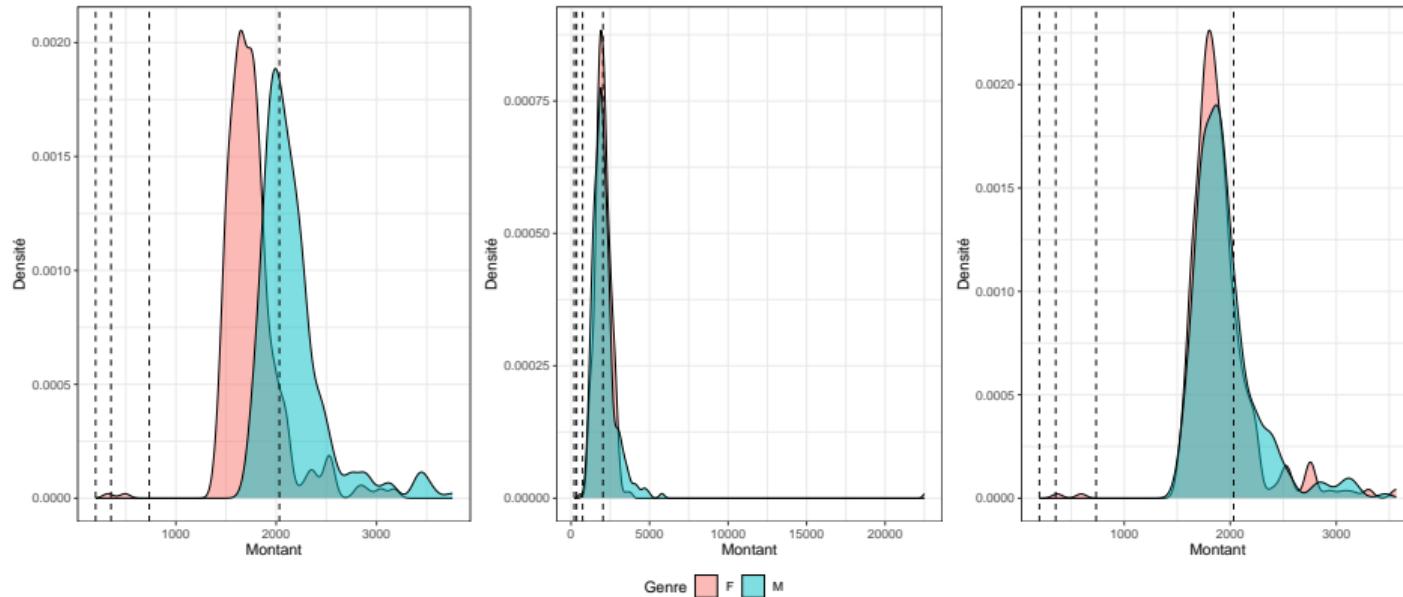


Figure – Densité des prédictions des modèles (excluant 0) en test

Gauche : Discrimination directe. Milieu : Modèle pénalisé. Droite : Discrimination indirecte (*fairness through unawareness*).

Conclusion

En résumé

- Définition d'une nouvelle métrique PAQ
- Modifier la fonction de perte
- Baisse de 84% de la PAQ en modifiant la fonction de perte

Améliorations et discussion

Points forts

- Le code permet de gérer plusieurs attributs catégoriels protégés
- Code modulable : il est aisément d'ajouter d'autres distributions (Tweedie) et mesure d'équités
- Permet de privilégier l'équité selon λ choisi
- Considérer les attributs protégés continus : discréteriser

Améliorations possibles

- Regarder l'effet de l'utilisation de quantiles différents sur la discrimination du modèle
- Explorer l'effet de la corrélation entre des variables et l'attribut protégé sur la pénalisation

En pratique...

Comment utiliser ces résultats ?

Conjecture actulab

Le produit de deux prédictions équitables est généralement équitable.

- Permet d'étudier séparément la fréquence/présence (Y) et la sévérité (M)
- **Sembla** naturel, mais aucune démonstration dans le cadre de cet atelier d'innovation

Code



Le projet est disponible sur [GitHub](#).



Références |

- [1] Dolman, C. et D. Semenovich. 2018, «Algorithmic fairness : Contemporary ideas in the insurance context», .
- [2] MS.UT. 2019, «Generalized linear models : Models with continuous variables», https://courses.ms.ut.ee/MTMS.01.011/2019_spring/uploads/Main/GLM_slides_5_continuous_response_print.pdf. Accessed : 2022-11-15.