

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. Démarche proposée
3. Expériences
4. Conclusion

Problématique

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 aimerais 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

- 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

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.

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 ($\mathcal{D}(\cdot)$).

Conjecture actulab

Conjecture actulab

Le produit de deux prédictions équitables est généralement équitable

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

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 ?

Intuition pour le cas continu

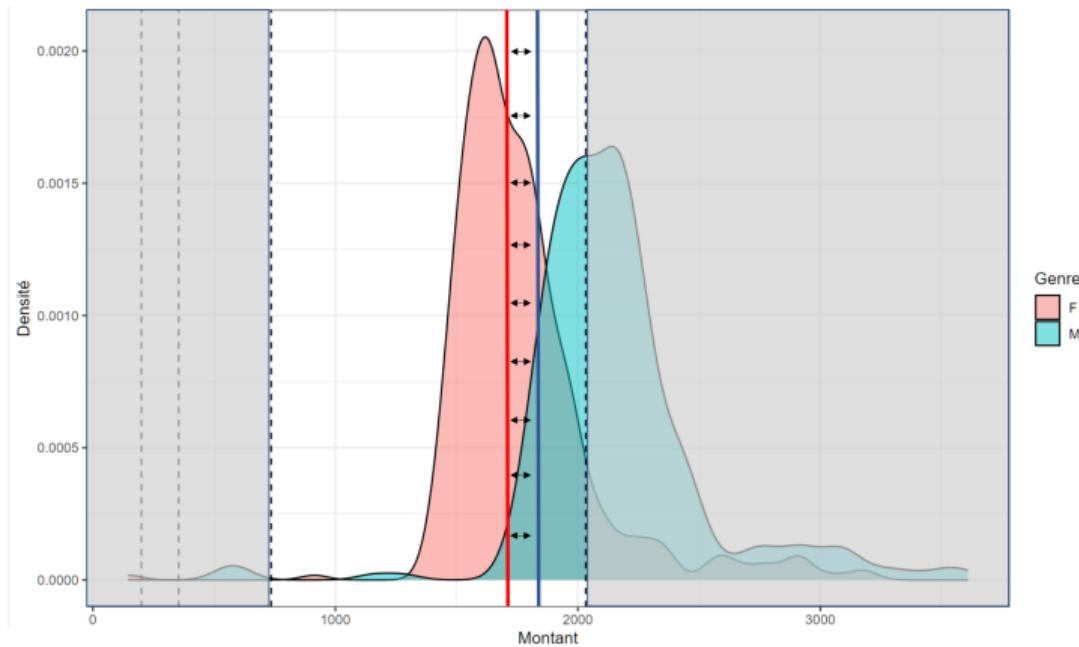


Figure – Écart entre les espérances des montant réclamés pour les hommes et les femmes entre deux quartiles

Parité Actulab par quartile

PAQ (Adaptée de Dolman et Semenovich [1])

$$\begin{aligned} & \left| \mathbb{E} \left[\text{montant prédit} \mid Q_\alpha \leq \text{montant réclamé} < Q_{\alpha+\frac{1}{4}}, \text{homme} \right] - \right. \\ & \left. \mathbb{E} \left[\text{montant prédit} \mid Q_\alpha \leq \text{montant réclamé} < Q_{\alpha+\frac{1}{4}}, \text{femme} \right] \right| \\ & < \delta(Q_{\alpha+\frac{1}{4}} - Q_\alpha), \end{aligned} \tag{3}$$

où Q_α sont les quartiles des montants réclamés et $\delta \geq 0$.

Pénalisations utilisées

Equalized Odds (variable réponse binaire)

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

Pénalisations utilisées

Equalized Odds (variable réponse binaire)

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

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

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

Expériences

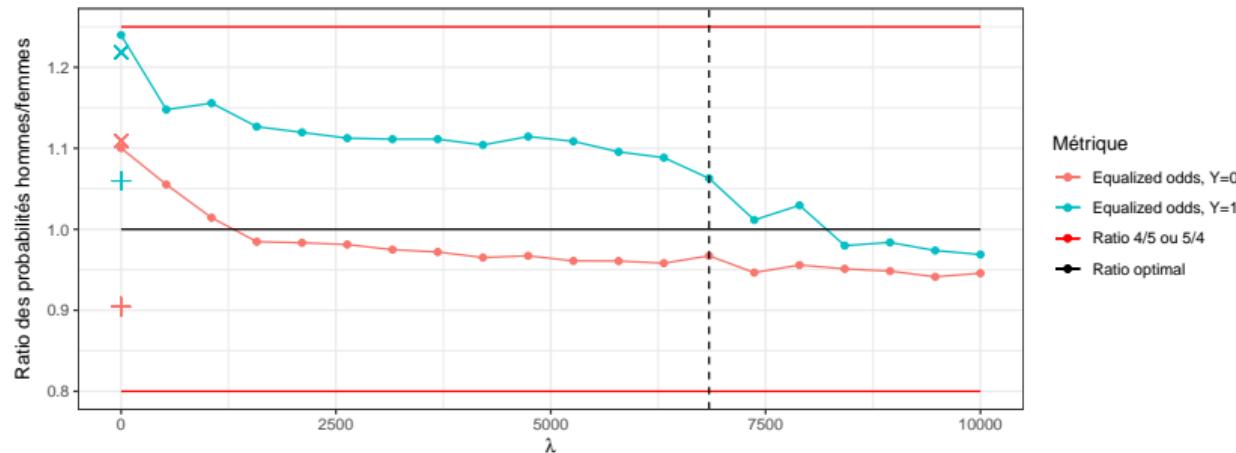
Modèles comparés

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

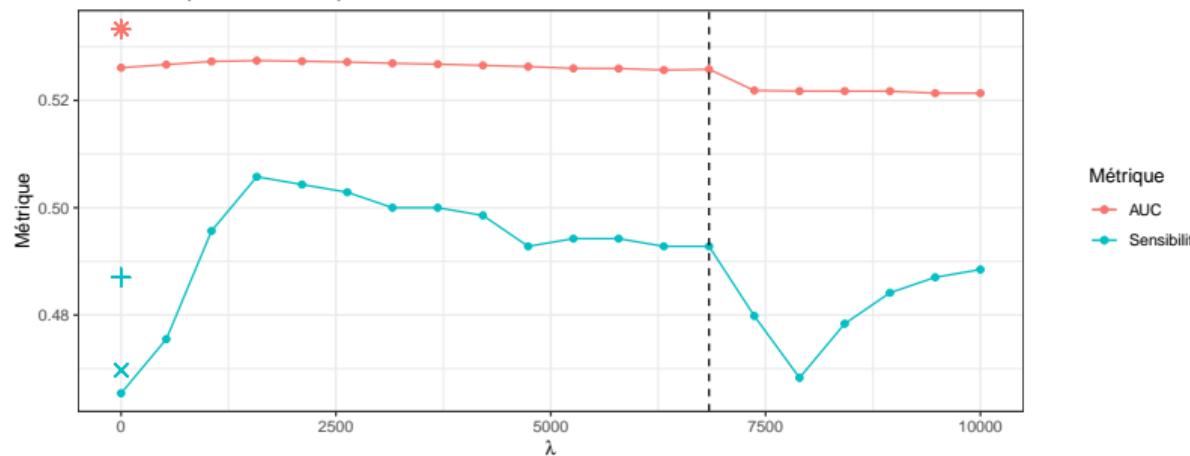
Performance prédictives des modèles

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

Effet de la pénalité sur l'équité



Effet de la pénalité sur la performance



Intuition de la métrique

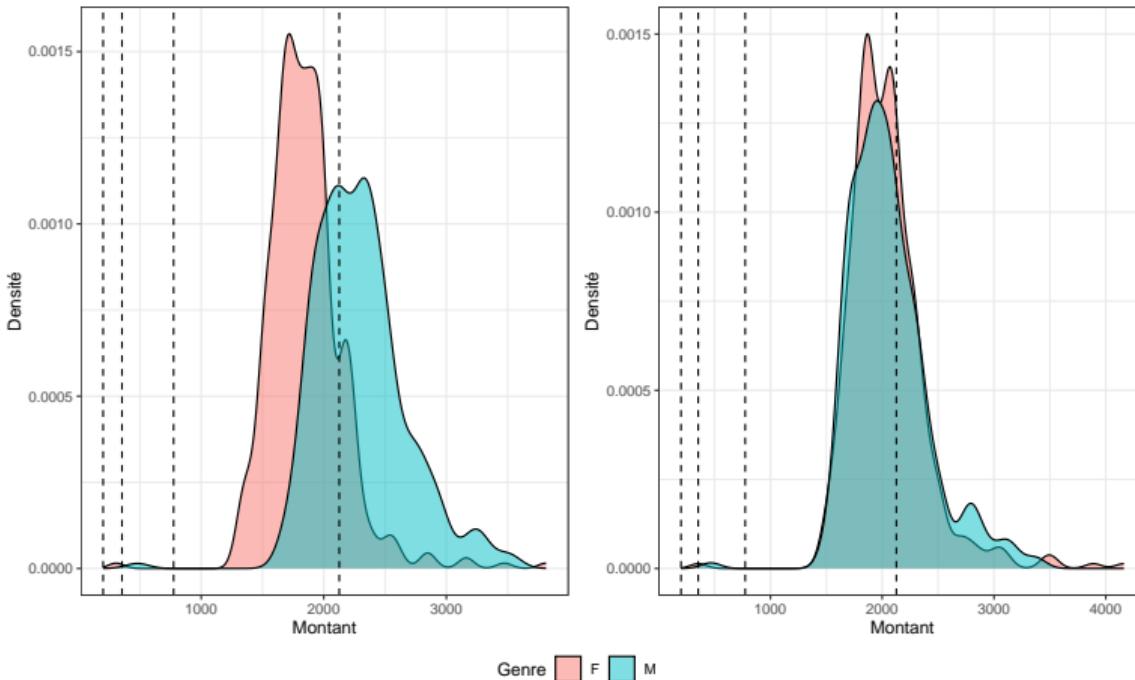


Figure – Répartition des prédictions (excluant 0) homme et femme pour la discrimination directe (gauche) et indirecte (droite)

Intuition de la métrique (suite)

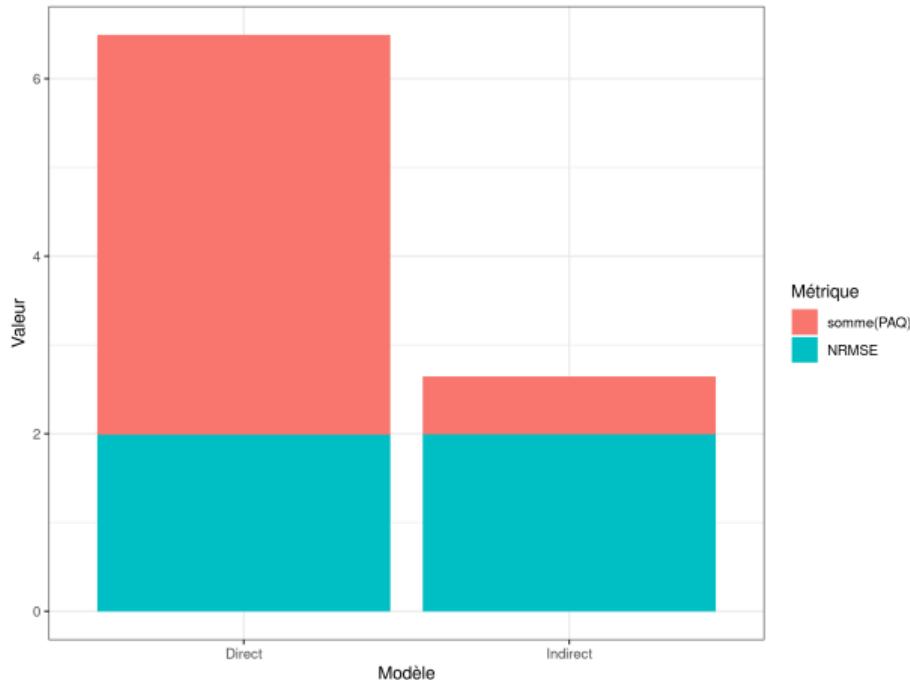
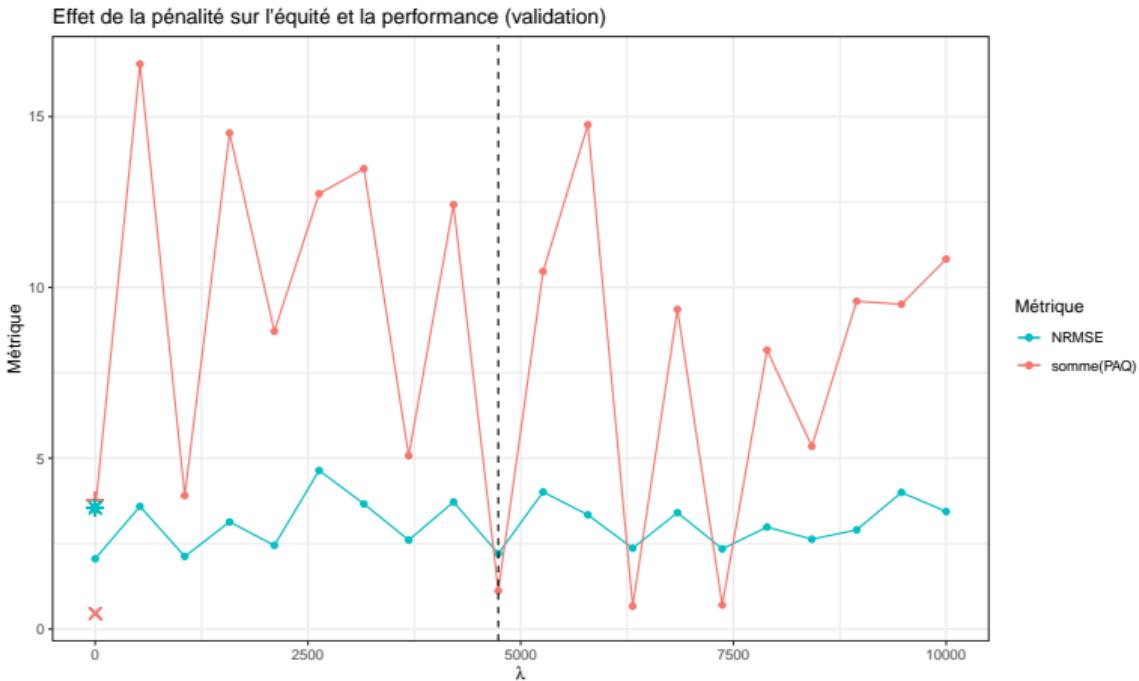


Figure – NRMSE et PAQ pour le modèle de discrimination directe et indirecte (excluant 0)

Résultats



+ : discr.
directe
x : discr.
indirecte

Résultats (suite)

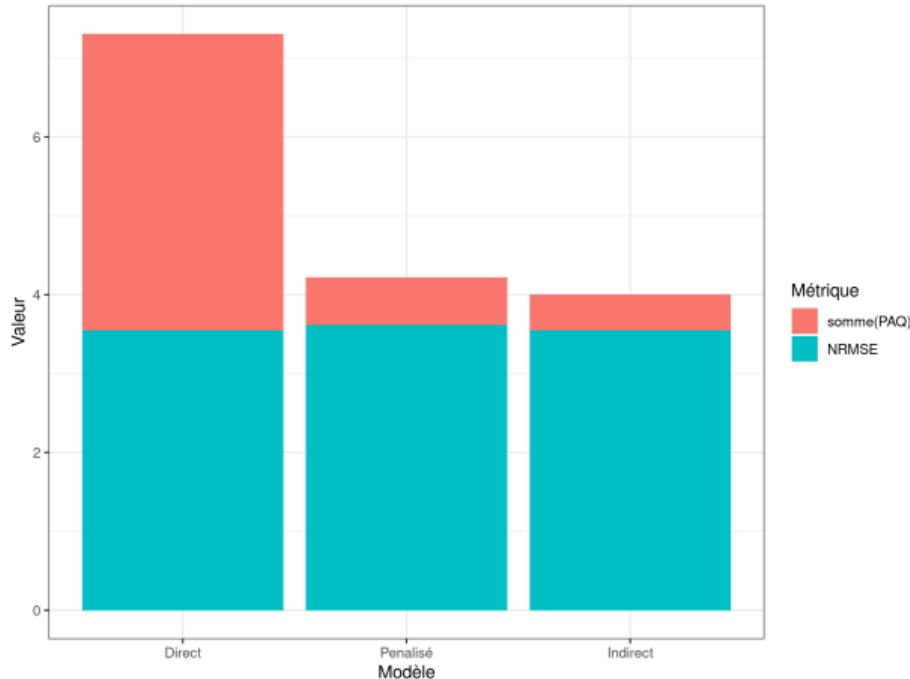


Figure – NRMSE et PAQ pour le modèle de discrimination directe, indirecte et modèle pénalisé (test)

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

- Valider la conjecture
- Regarder l'impact de l'utilisation de quantiles différents sur la discrimination du modèle

Code



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



Références I

- [1] Dolman, C. et D. Semenovich. 2018, «Algorithmic fairness : Contemporary ideas in the insurance context», .