MESURE ET MITIGATION DES BIAIS : VERS UNE TARIFICATION NON-VIE RÉELLEMENT ÉQUITABLE

Mulah MORIAH
EURIA 2022

10/10/2023

Sommaire et objectifs

A la fin de cette présentation, vous comprendrez tous les tenants et aboutissants des sujets d'équité et leur impact sur les différentes contraintes en tarification. Vous saurez parler des aspects techniques, les vulgariser et vous apprendrez à implémenter l'équité dans un contexte de tarification réel.

Assurance IARD et tarification

Introduire le sujet de l'équité en partant d'éléments connus en IARD.

Présenter l'intérêt et la place de l'équité.

01

Définition de l'équité

Comprendre ce que représente l'équité littérairement et théoriquement.

Découvrir les origines des discriminations.

02

Mesures et mitigations

Etudier les mesures d'équité.

Acquérir une connaissance des outils mathématiques, de leur limite et leur impact en tarification.

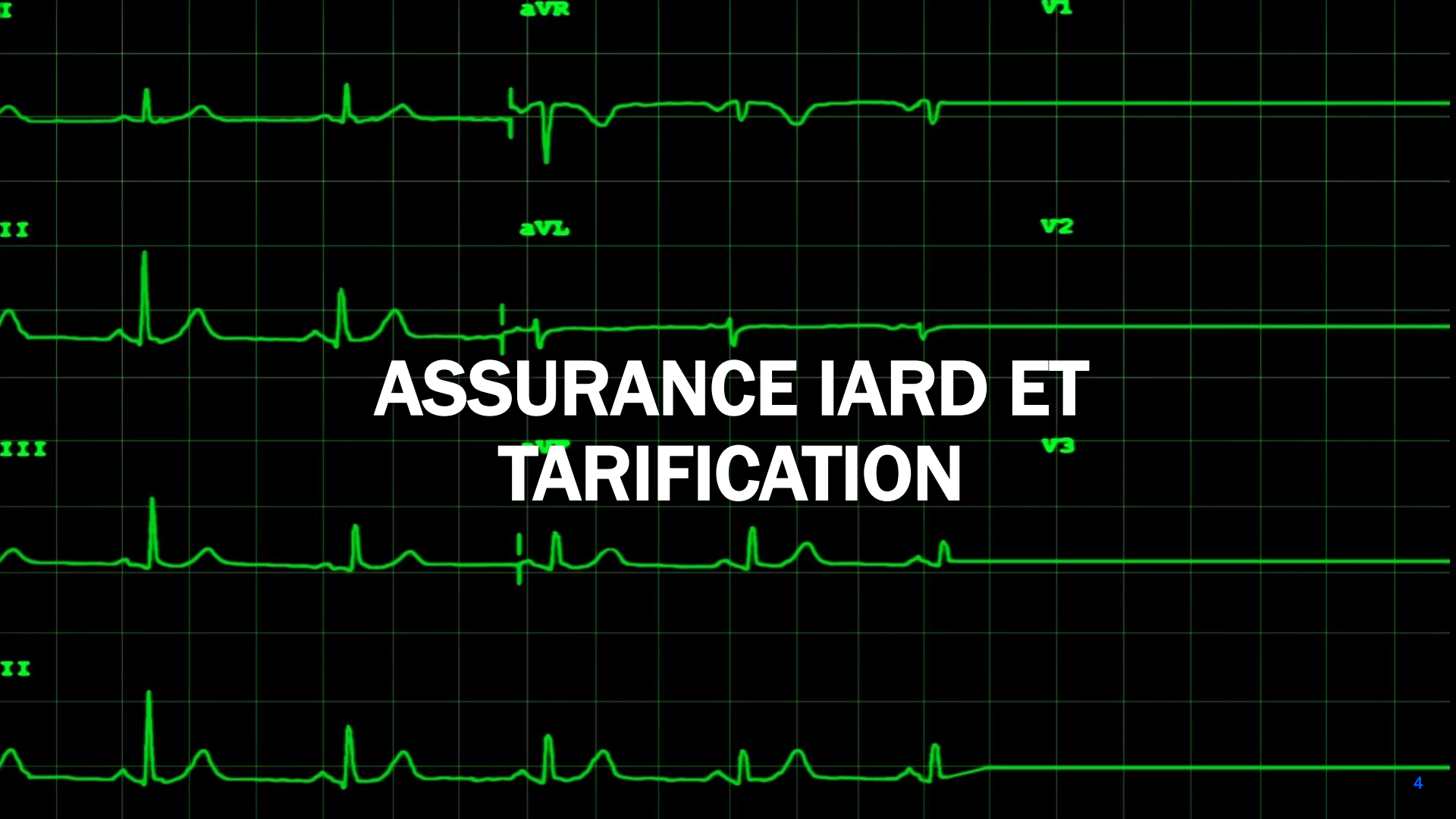
03

Mise en application

Comprendre comment réduire théoriquement les discriminations.

Etudier les détails de la mise en place de l'équité en pratique.

04



ASSURANCE IARD ET TARIFICATION

Les contraintes de tarification

La prime comme tout prix encapsule de nombreux enjeux stratégiques.

Qualité et cohérence statistique

L'inversion du cycle de production

Contraintes commerciales

Tarification concurrentielle

Contraintes règlementaires

Gender directive

Contraintes stratégiques

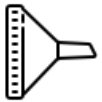
Favoriser certaines zones ou certaines clientèles

La “discrimination” actuarielle comme paradigme de tarification

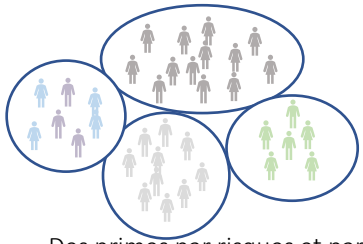
La tarification une forme de "discrimination" basée sur une représentation du risque porté par l’assuré.

Des assurés définis par les informations dont dispose l’assureur:

- Informations personnelles de l’assuré
- Informations sur son bien
- Informations sur son contrat
- Informations sur ses sinistres



Méthodes statistiques effectuant des formes de segmentation suivant les informations à la disposition de l’assureur.



Des primes par risques et par segments.

Quelle est la frontière entre discrimination acceptable et inacceptable pour la tarification ?

La gender directive

L'exemple le plus connu est celui de la gender directive.

Les primes finales affichées par les assureurs ne doivent pas faire de distinction entre les genres.

Les solutions implémentées :

- Supprimer le genre des modèles
- Retraiter le genre en sortie des modèles

Pourquoi donc remettre le sujet des discriminations sur la table ?!

Les discriminations sont toujours présentes

Ces résultats se formalisent :

Notre tableau :

$n_{i,j}$	Femme	Homme	$n_{i,\bullet}$
A	45	8	53
B	20	53	73
$n_{\bullet,j}$	65	61	126

$e_{i,j}$	Femme	Homme	$e_{i,\bullet}$
A	151	33	184
B	129	290	419
$e_{\bullet,j}$	280	323	603

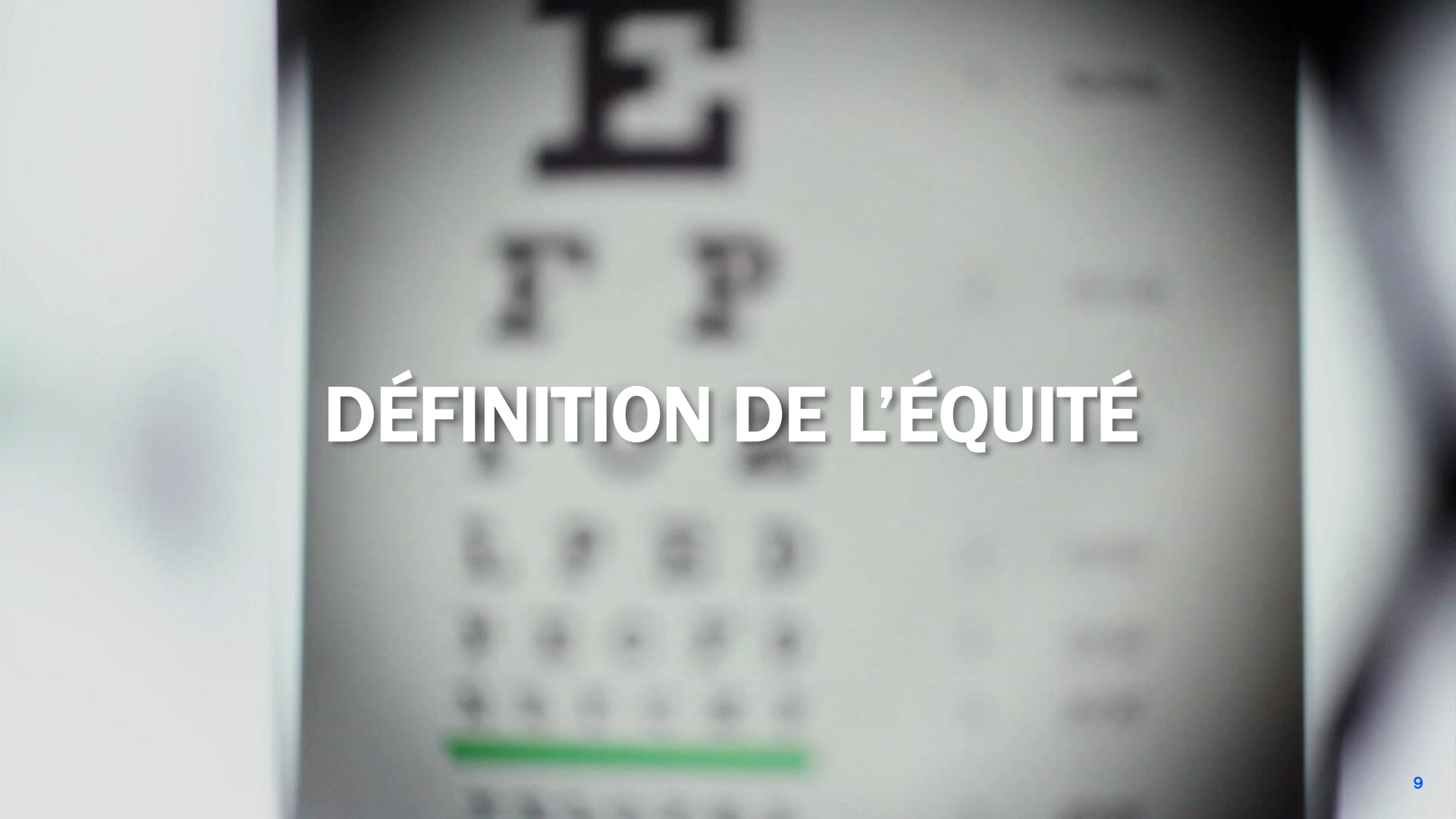
Nos estimations :

$$\widehat{\mu}_{1,\bullet} = \frac{n_{1,\bullet}}{e_{1,\bullet}} = \frac{53}{184} = 0.288$$

$$\widehat{\mu}_{2,\bullet} = \frac{n_{2,\bullet}}{e_{2,\bullet}} = \frac{73}{419} = 0.174$$

La formalisation de l'effet indirecte du genre :

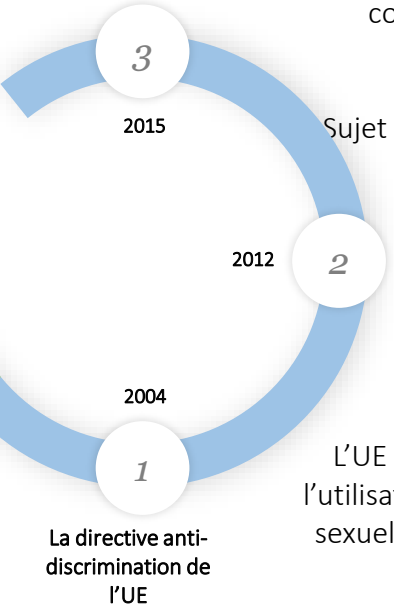
$$\widehat{\mu}_{1,\bullet} = \widehat{\mu}_{1,1}\widehat{\mathbb{P}}(\text{Femme}|A) + \widehat{\mu}_{1,0}\widehat{\mathbb{P}}(\text{Homme}|A) = \widehat{\mu}_{1,1}\frac{e_{1,1}}{e_{1,1}+e_{1,0}} + \widehat{\mu}_{1,0}\frac{e_{1,0}}{e_{1,1}+e_{1,0}}$$



DÉFINITION DE L'ÉQUITÉ

L'équité : entre abstraction et réalité

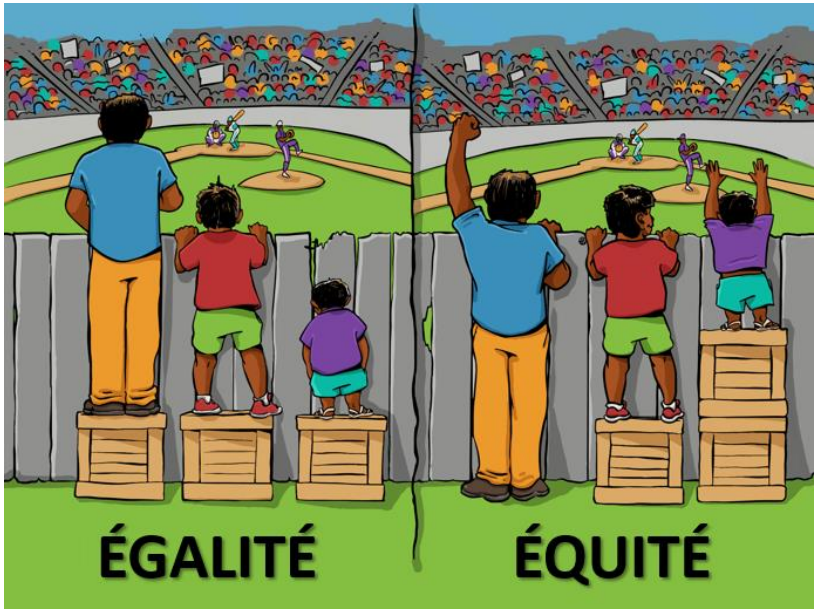
- Le droit à l'oubli
- Suppression de formulaire médicale
- Solidarité générationnelle



Donner les mêmes opportunités à chacun en prenant en compte les différences qui existent entre chaque individu.

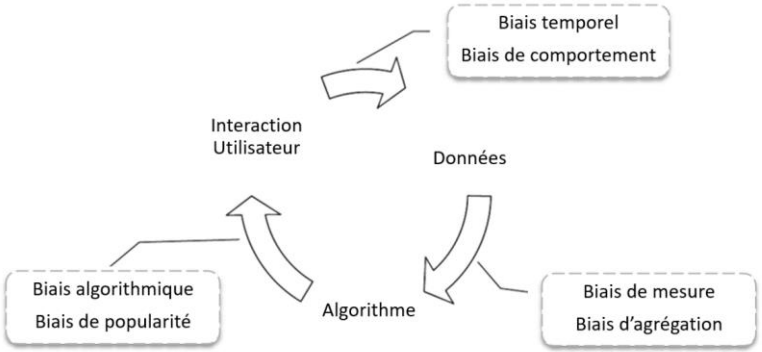
Sujet évoqué depuis -350AV, en droit et en philosophie.

L'UE parle d'équité pour proposer le bannissement de l'utilisation de variables dites sensibles : genre, orientation sexuelle, handicap, maladies graves ou héréditaires dans des secteurs comme l'assurance.

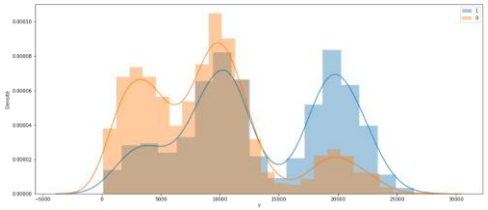


Inspirée de l'encyclopédie canadienne

Origines et subtilités des discriminations



Biais historique
Exemple de google images 2016



Biais statistique
Exemple des distributions par genre

Biais historique : des faits statistiques ? Les biais historiques ont la particularité de devenir dans certains cas des faits statistiques. Ils ne sont plus considérés comme des biais mais sont considérés à tort comme des faits observés et justifiables. Le cas du genre est assez remarquable. Des décennies d'histoire ont créé des biais liés au genre. Ces biais sont si ancrés dans l'histoire qu'ils sont difficiles à dissocier du reste de la réalité. Il y a les exemples classiques des carrières professionnelles, de l'accès aux responsabilités

Le piège des faits statistiquement justifiables
Exemple de UBER

MESURES ET MITIGATIONS

Le challenge de la définition des biais

Une littérature récente à la recherche de consensus

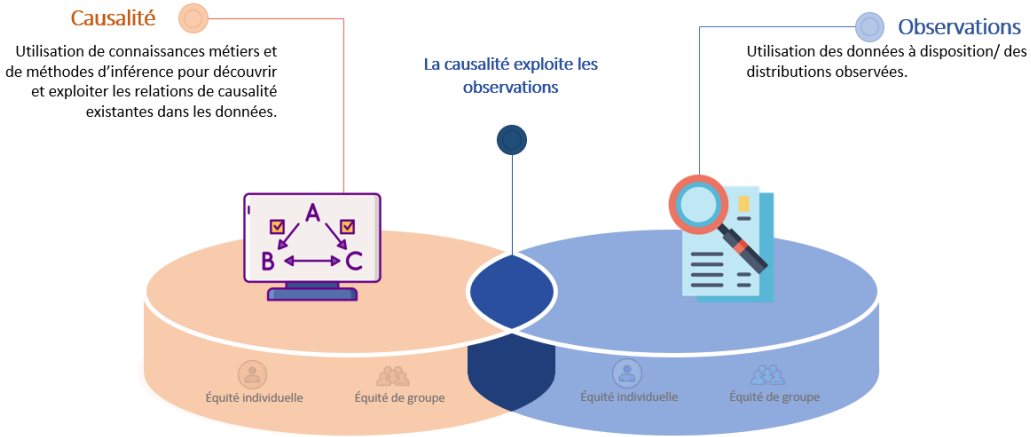
- Début des intérêts 2016-2018

- De nombreuses définitions et approches différentes, des auteurs parlent du « zoo » des définitions
- Les premiers travaux traitant de l'assurance publiés début 2022
- Littérature principalement anglophone et axée sur la classification

D'où l'utilité d'une revue de la littérature dans le but de poser un cadre clair et complet, mais aussi axée sur la tarification et la régression.

Variable sensible : S
Variable d'intérêt : Y
Variable d'intérêt prédite : $\hat{Y}$

Les différentes formes d'équité



Indépendance $\widehat{Y} \perp\!\!\!\perp S$

Exemple : Parité statistique

$$\mathbb{P}(\widehat{Y} \leq y|S) = \mathbb{P}(\widehat{Y} \leq y)$$

Séparation $Y \perp\!\!\!\perp S|\widehat{Y}$

Exemple : chances égalisées

$$\mathbb{P}(\widehat{Y} \leq y|S, Y) = \mathbb{P}(\widehat{Y} \leq y|Y)$$

Suffisance $\widehat{Y} \perp\!\!\!\perp S|Y$

Exemple : chances égalisées

$$\mathbb{P}(Y \leq y|S, \widehat{Y}) = \mathbb{P}(Y \leq y|\widehat{Y})$$

Distance individuelle $d_Y(\widehat{y}_i, \widehat{y}_j) < \lambda d_X(x_i, x_j)$

Exemple : distance de Dwork

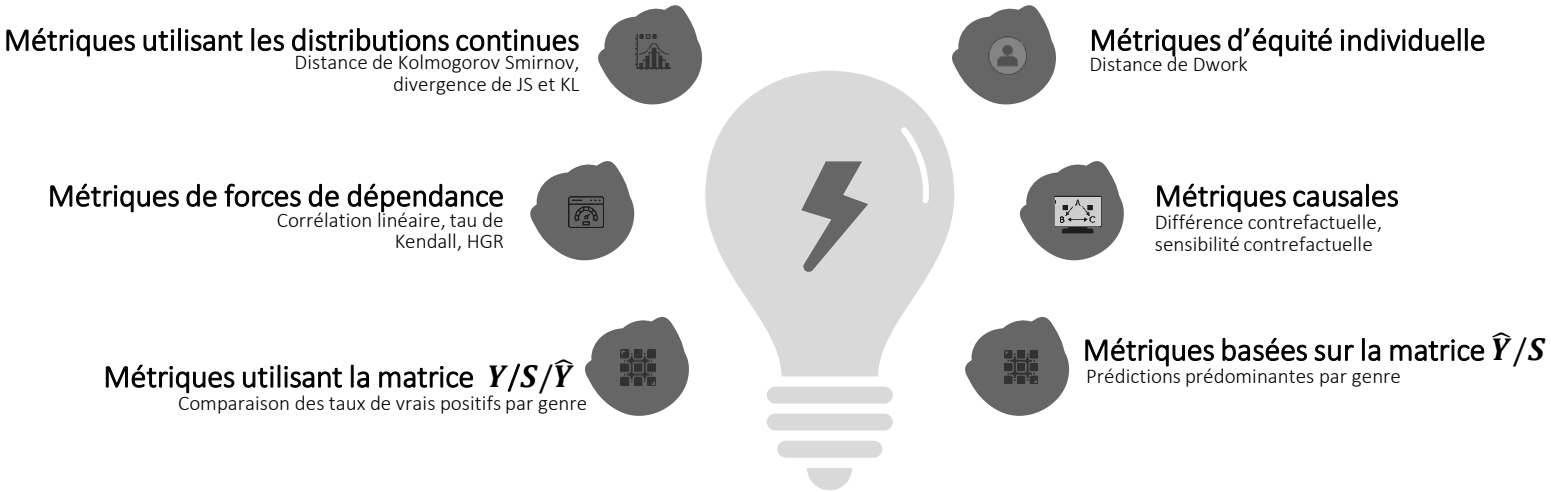
$$d_X = 1 - \frac{1}{n}\left(\sum_{i=1}^{n}\left|\widehat{y}_i - \frac{1}{k}\sum_{x_j \in \mathbb{V}_{\mathrm{KNN}}(x_i)} \widehat{y}_j\right|\right)$$

Contre-factualité

Exemple : espérance (do-calculus

$$\mathbb{P}(\widehat{Y}_{S\leftarrow 1} = 1|S = 1) = \mathbb{P}(\widehat{Y}_{S\leftarrow 0} = 1|S = 1)$$

Métriques d'équité



Les métriques de biais : HGR

Permet de mesurer **toutes formes de dépendances** et respecte les sept propriétés fondamentales d'une bonne mesure de dépendance énoncées par Rényi en 1959.

$$HGR(Y,S) = \sup_{f:\mathcal{Y}\mapsto\mathbb{R},\ g:\mathcal{S}\mapsto\mathbb{R}} \rho(f(Y),g(S)) = \sup_{f:\mathcal{Y}\mapsto\mathbb{R},\ g:\mathcal{S}\mapsto\mathbb{R}} \mathbb{E}(f(Y),g(S))$$

Valeur exacte difficile, voire impossible à calculer pour l'instant.

Mary et al. fournissent un estimateur, appelée HGR KDE, permettant d'approcher le HGR.

Cet estimateur permet de maintenir les propriétés du HGR et a prouvé empiriquement ses performances.

En absence de biais, la dépendance doit être nulle.

Les métriques de biais : adaptation flip-test

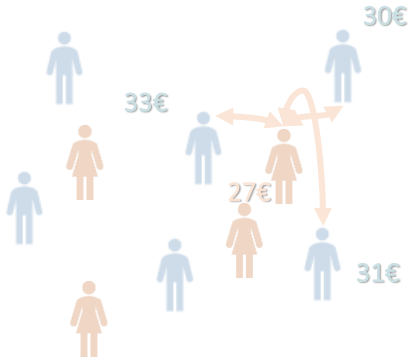
La méthode flip-test initialement construite pour le cas de la classification est adaptée pour répondre à la question : quelle serait la prime d'une femme si elle était un homme ?

Pour cela :

1. L'algorithme de k plus proches voisins permet de définir les individus semblables en se servant de toutes les informations sauf la variable sensible et la variable d'intérêt.

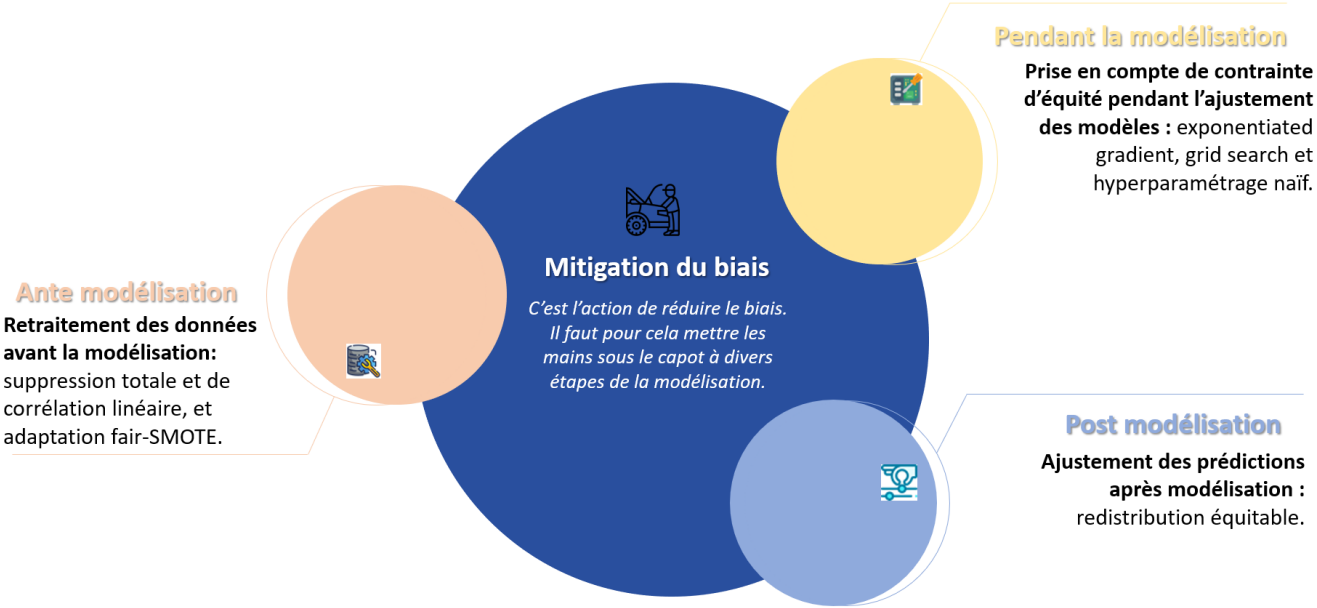
2. Une fois les voisinages construits, les écarts de primes entre individus et individus du genre opposé sont calculés.

En absence de biais, ces écarts doivent être nuls.



Ecart femme i = 27€ - (31€+30€+33€)/3
= - 4,3€

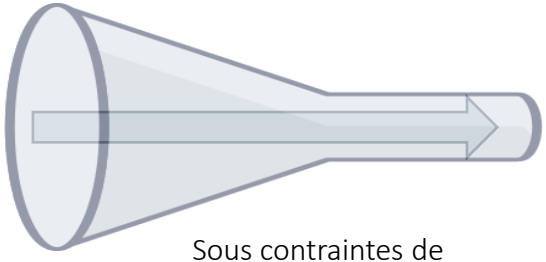
Les approches de mitigation du biais



Ante modélisation: exemple de la suppression

Retraitement des données, traitements des valeurs manquantes, construction de nouvelles variables, suppression de variables

Données brutes



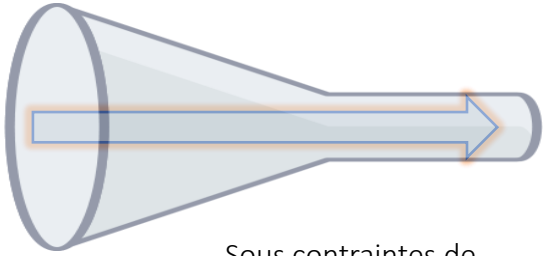
Données de modélisation

Sous contraintes de performance et de cohérence

Ante modélisation: exemple de la suppression

Retraitement des données, traitements des valeurs manquantes, construction de nouvelles variables, suppression de variables

Données brutes



Données de modélisation

Sous contraintes de performance, de cohérence et d'équité

Prise en compte de l'équité

Par exemple, sélectionner les variables en prenant en compte l'équité.

Avantages

- ✓ Conservation de l'ensemble du processus de modélisation
- ✓ Simple et moins coûteuse

Inconvénients

- Perte d'informations
- Réinjection de biais dans le reste du processus

Pendant modélisation: exemple de l'exponentiated gradient

La construction d'une fonction de prédiction peut se résumer à la résolution d'un programme de minimisation: minimiser l'erreur de prédiction.

$$\arg\min_{f \in \mathbb{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(X_i)) \right\}$$

Pendant modélisation: exemple de l'exponentiated gradient

La construction d'une fonction de prédiction peut se résumer à la résolution d'un programme de minimisation: minimiser l'erreur de prédiction.

$$\arg\min_{f \in \mathbb{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(X_i)) \right\}$$

Une contrainte d'équité peut être rajoutée à ce système de minimisation.

La méthode exponentiated gradient, issue de la théorie des jeux, permet de trouver le meilleur arbitrage entre minimisation de l'erreur et maximisation de l'équité.

Avantages

- ✓ Utilise le plus d'informations sur S
- ✓ Permette en théorie le meilleur arbitrage entre équité et performance

Inconvénients

- \- Difficile à implémenter et à généraliser
- \- Coûteux en temps de calculs et faire converger

Post modélisation: exemple de la redistribution équitable

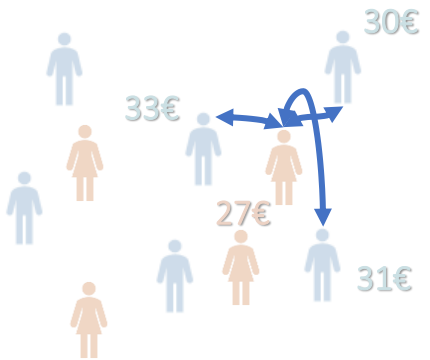
Idée de base : déplacer la frontière de décision en prenant en compte des contraintes d'équité

Régression logistique : déplacer le seuil ou faire des seuils différents par genre.

- Pas de frontières de décision en régression
- La prédiction avantageuse ne peut pas être définie dans l'absolu

Solution : « redistribution équitable »

- Définir à l'aide de l'adaptation flip-test la qualité de la prime de chaque individu
- Corriger les primes pour les rapprocher des primes des individus du genre opposé



Ecart femme i = 27€ - (31€+30€+33€)/3
= - 4,3€

Post modélisation: exemple de la redistribution équitable

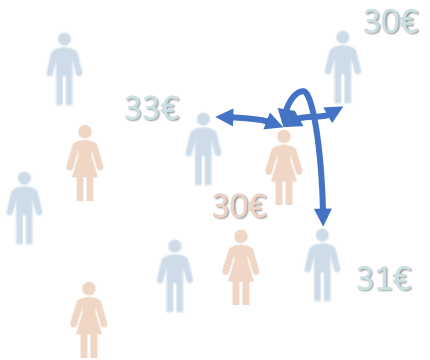
Idée de base : déplacer la frontière de décision en prenant en compte des contraintes d'équité

Régression logistique : déplacer le seuil ou faire des seuils différents par genre.

- Pas de frontières de décision en régression
- La prédiction avantageuse ne peut pas être définie dans l'absolu

Solution : « redistribution équitable »

- Définir à l'aide de l'adaptation flip-test la qualité de la prime de chaque individu
- Corriger les primes pour les rapprocher des primes des individus du genre opposé



Ecart femme i = 30€ - (31€+30€+33€)/3
= - 1,3€

Post modélisation: exemple de la redistribution équitable

Redistribuer récursivement des parties des écarts tout en contrôlant la qualité de la distribution obtenue, le niveau de biais mitigé et le S/P.

Deux hyperparamètres sont utilisés pour contrôler la vitesse de correction et le niveau minimum de biais à atteindre.

Tant que biais > seuil :

- Corriger prime i de écart/vitesse de correction
- Remesurer le biais de chaque individu
- Alterner le genre

Avantages

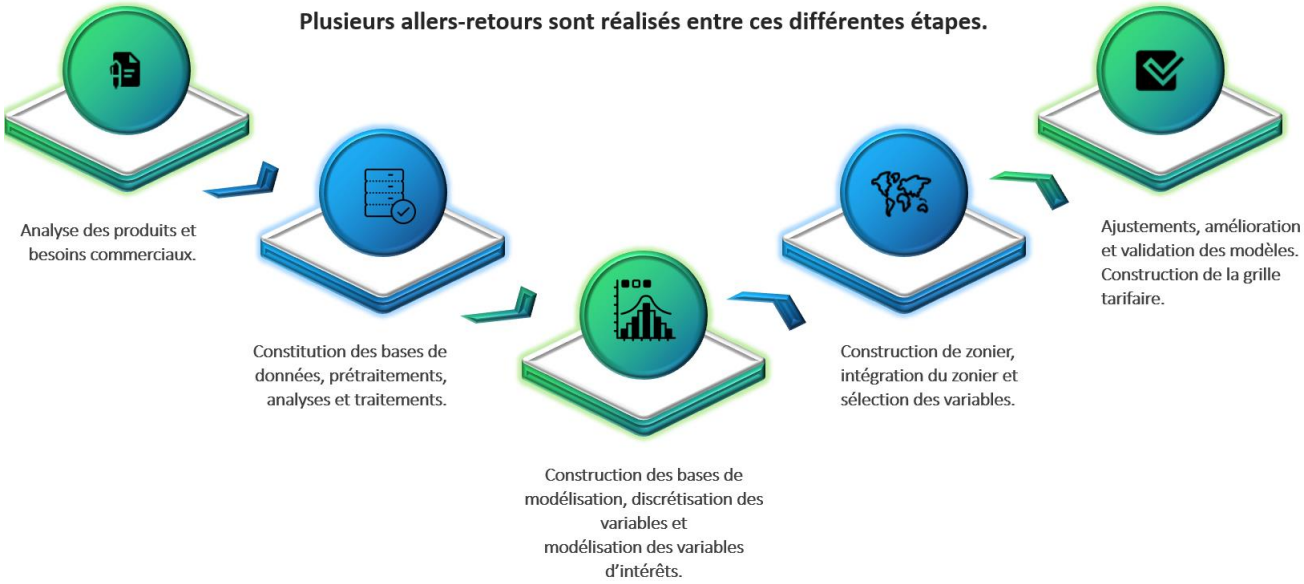
- ✓ Pas de recalibrages nécessaires
- ✓ Fourni une approche individuelle

Inconvénients

- \- Résultats dépendants de la qualité des modèles construits

Mise en application de l’équité

Périmètre de l'étude



Tarification classique

1. GLM
2. GRADIENT BOOSTING
3. FORET ALEATOIRE

Modèles	RMSE	MAE	True/Pred
GLM	171,08	52,94	1,010
GB	172,33	53,07	1,023
RF	171,04	52,94	1,017

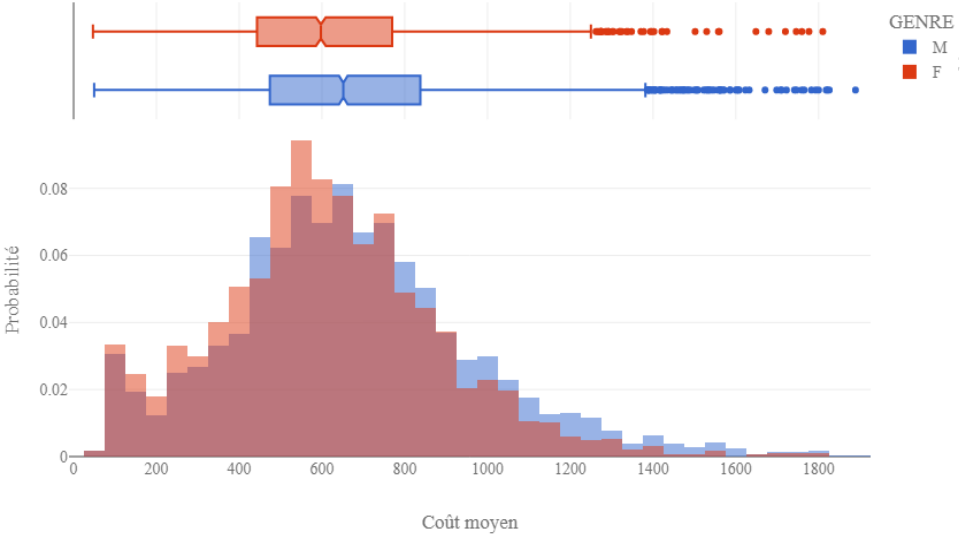
Performances des modèles de primes pures

RF: Le gain de précision < **GLM** : interprétabilité + place sur le marché

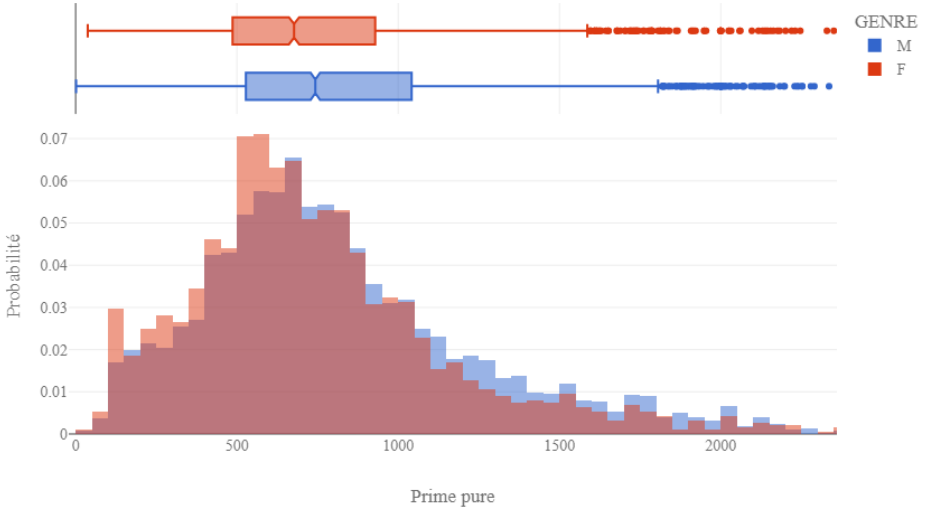
- Modélisation sans le genre
- GLM retenu comme référence
- Modèle de prime pure retenu à la place d'un modèle coût x fréquence

Mesures des biais : avant modélisation

Y	Kendall	HGR_KDE	Flip-test
Coût moyen	-5,9%	8,1%	-12,41€
Fréquence	-1,8%	31,6%	-2,14%
Prime	-1,8%	31,7%	-10,78€



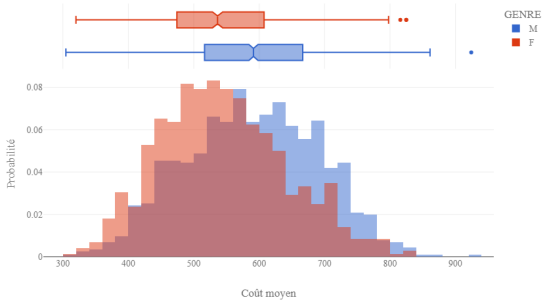
Distribution du coût moyen par genre.



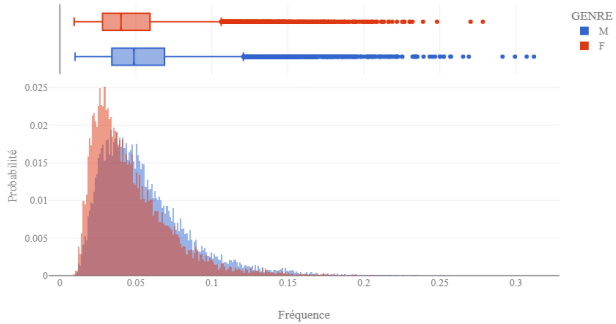
Distribution de la prime pure par genre.
« Prime pure » individuelle, poids en 0 retiré.

Mesures des biais : après modélisation avec le genre

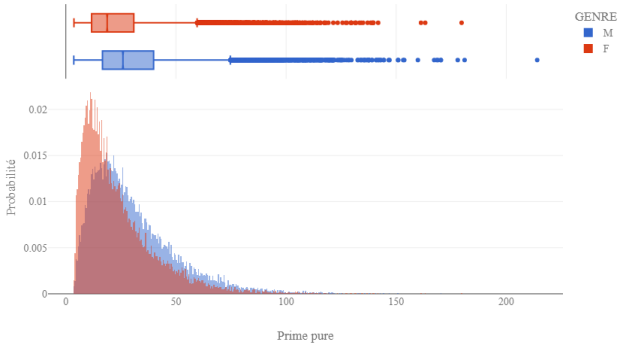
Y	Kendall	HGR_KDE	Flip-test
Coût moyen	-18,2%	21,3%	-3,16€
Fréquence	-18,4%	31,9%	-1,12%
Prime	-20,2%	33,7%	-1,21€



Distribution du coût moyen par genre.



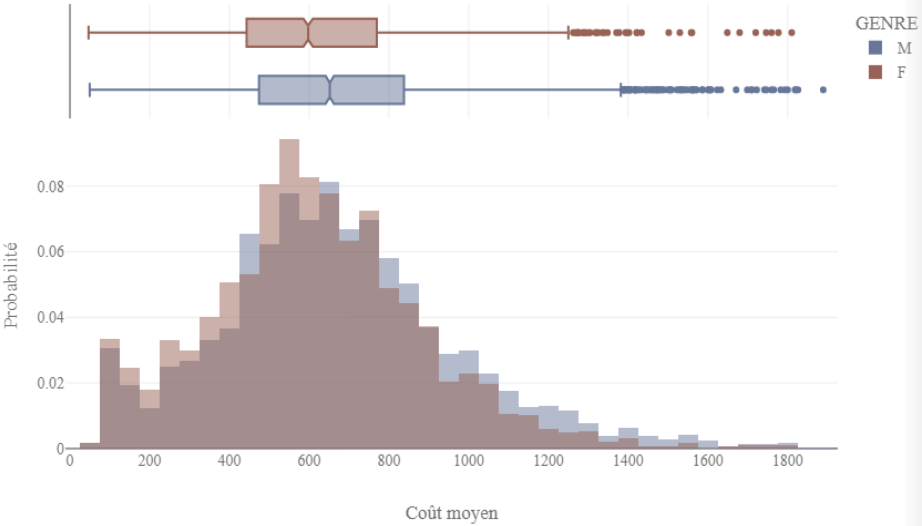
Distribution de la fréquence par genre.



Distribution de la prime pure par genre.

Mesures des biais : après modélisation avec le genre

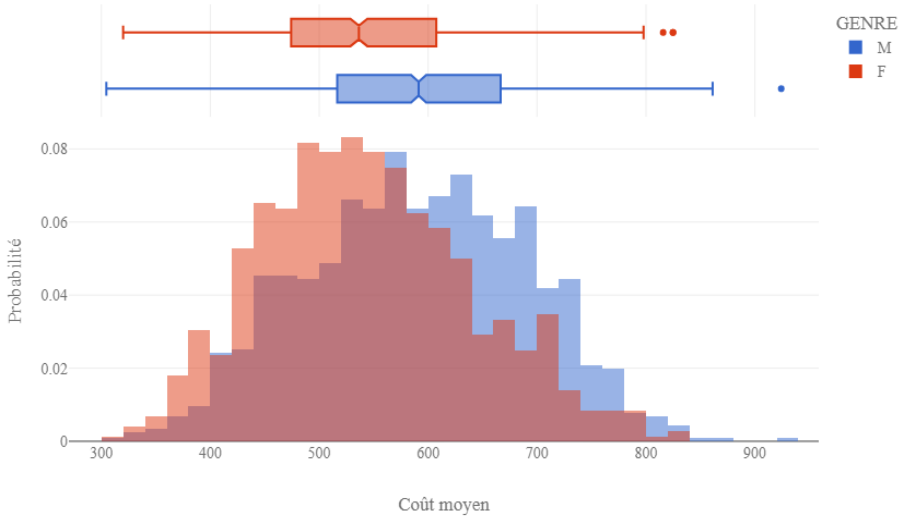
Y	Kendall	HGR_KDE	Flip-test
Coût moyen	-5,9%	8,1%	-12,41€
Fréquence	-1,8%	31,6%	-2,14%
Prime	-1,8%	31,7%	-10,78€



Distribution du coût moyen par genre avant modélisation

Avant

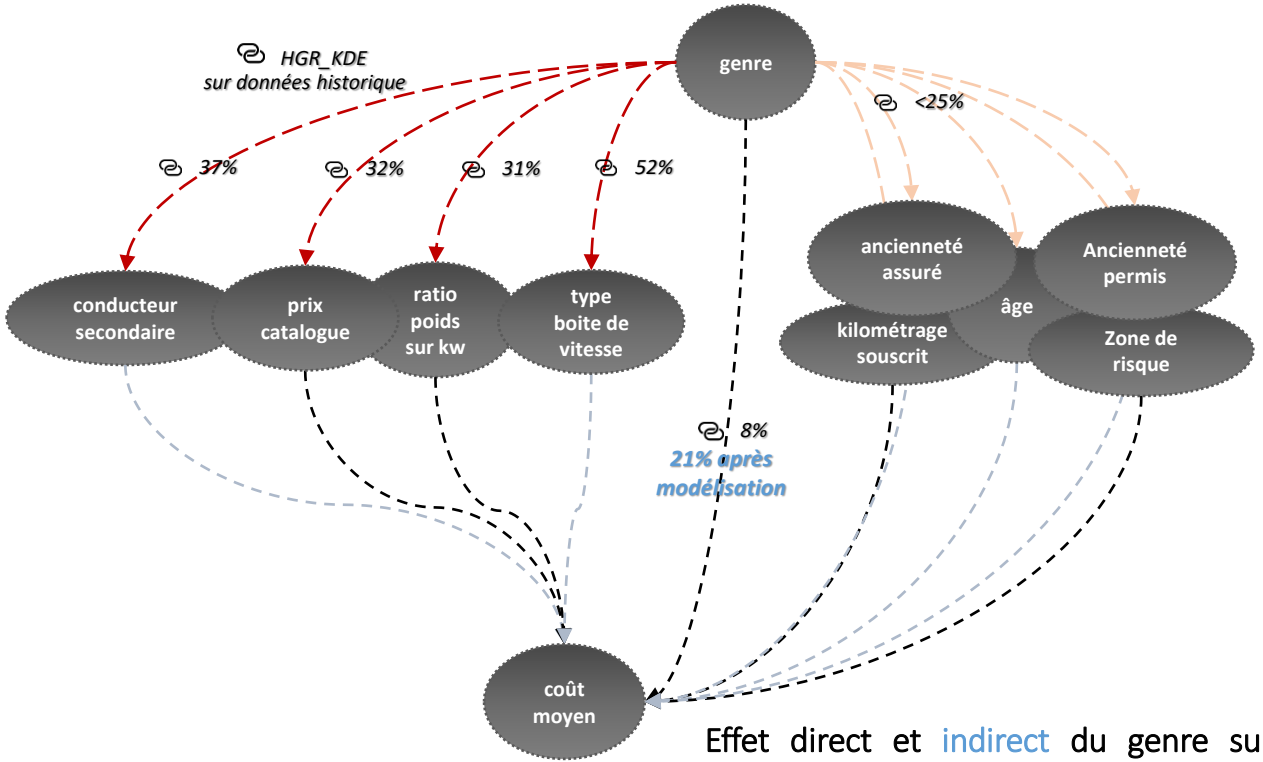
Y	Kendall	HGR_KDE	Flip-test
Coût moyen	-18,2%	21,3%	-3,16€
Fréquence	-18,4%	31,9%	-1,12%
Prime	-20,2%	33,7%	-1,21€



Distribution du coût moyen par genre après modélisation

Après

Comment expliquer l'amplification du biais ?



Effet direct et indirect du genre sur le coût moyen

Mesures des biais : après modélisation sans le genre

Institut des ACTUAIRES

Suppression du genre du modèle

Y	Kendall	HGR_KDE	Flip-test
Coût moyen	-17,3%	20,4%	-2,89€
Fréquence	-13,2%	28%	-1,03%
Prime	-17,4%	29,7%	-0,90€

Retraitement du genre en sortie du modèle

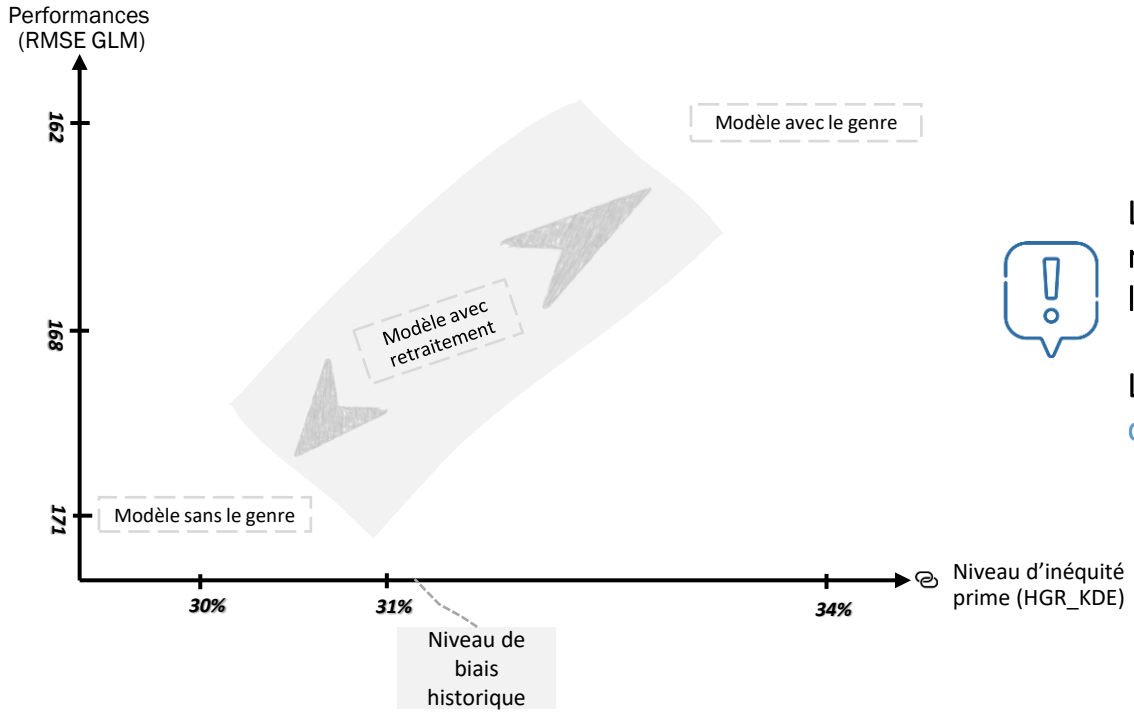
Y	Kendall	HGR_KDE	Flip-test
Coût moyen	-18,1%	20,9%	-2,33€
Fréquence	-13,1%	27,5	-0,93%
Prime	-17,2%	30,6%	-0,81€

Les résultats sont quasiment les mêmes qu'en présence du genre dans le modèle…

Y	Kendall	HGR_KDE	Flip-test
Coût moyen	-18,2%	21,3%	-3,16€
Fréquence	-18,4%	31,9%	-1,12%
Prime	-20,2%	33,7%	-1,21€

Les autres variables explicatives permettent de reconstruire le genre, et donc de maintenir son effet dans les différents modèles.

Mesures des biais : récapitulatif



La suppression du genre ne permet pas de réduire le biais.

Le genre a toujours autant d'effet sur les modèles.

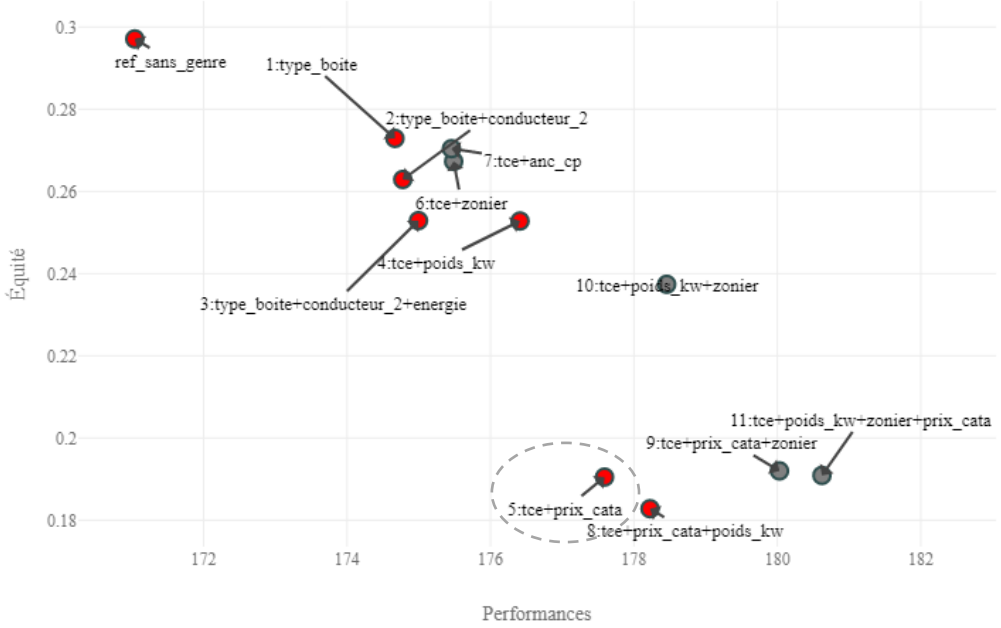
Mitigation du biais : suppression totale

Dépendance entre le genre et les autres variables.

❌ Des scénarios de suppression sont construits:

- ***Scénario 1*** *: type_boite ;*
- ***Scénario 2*** *: type_boite et conducteur_2 ;*
- ***Scénario 3*** *: type_boite, conducteur_2 et energie ;*
- ***Scénario 4*** *: type_boite, conducteur_2, energie et poids_kw ;*
- ***Scénario 5*** *: type_boite, conducteur_2, energie et prix_cata ;*
- ***Scénario 6*** *: type_boite, conducteur_2, energie et zonier ;*
- ***Scénario 7*** *: type_boite, conducteur_2, energie et anc_cp ;*
- ***Scénario 8*** *: type_boite, conducteur_2, energie, prix_cata et poids_kw;*
- ***Scénario 9*** *: type_boite, conducteur_2, energie, prix_cata et zonier ;*
- ***Scénario 10*** *: type_boite, conducteur_2, energie, poids_kw et zonier ;*
- ***Scénario 11*** *: type_boite, conducteur_2, energie, prix_cata, poids_kw et zonier.*

Mitigation du biais : suppression totale

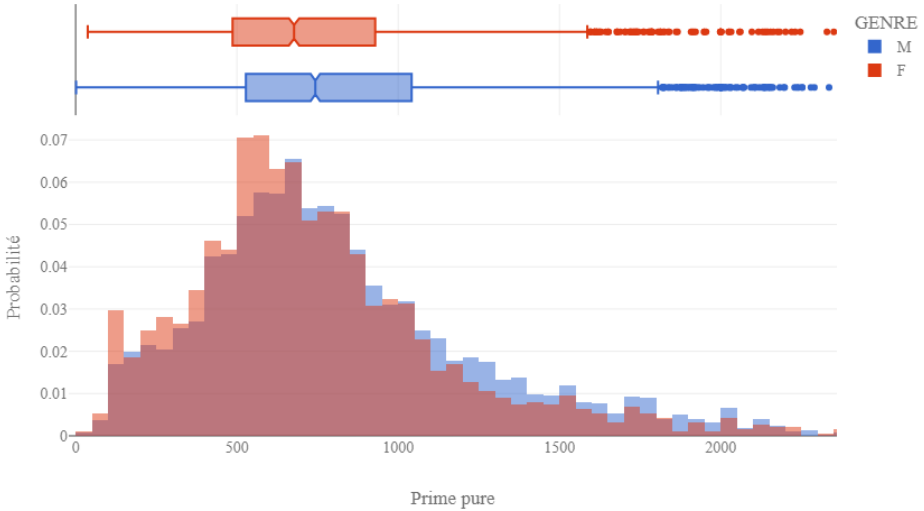


Scénario 5 : type boite, conducteur secondaire, énergie et prix catalogue

Indicateurs/Modèles	Référence	Suppression
RMSE	171,04	177,6
HGR_KDE	29,71%	19%
S/P	99,66%	99,30%

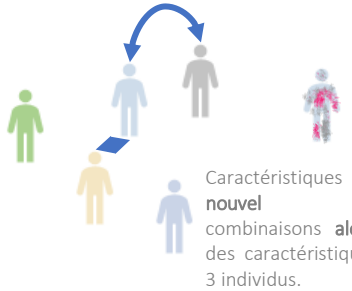
- Equité accrue pour une perte de performance acceptable
- Dépend de la capacité des variables restantes à remplacer les variables supprimées

Mitigation du biais : adaptation fair-SMOTE



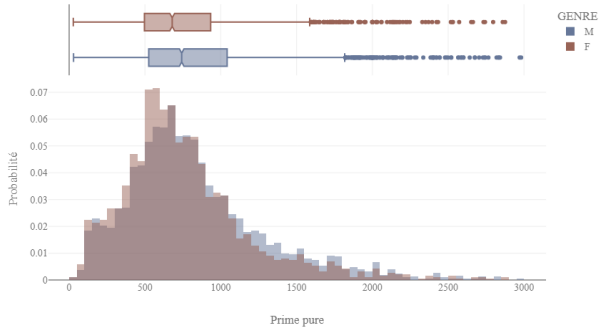
Distribution de la prime pure par genre.
« Prime pure » individuelle, poids en 0 retiré.

- Discrétisation de la prime en 7 zones
- « Rééchantillonnage » par une approche SMOTE utilisant les k plus proches voisins pour construire de nouveaux individus. Deux hyperparamètres contrôlent ce processus.
- Rééchantillonner pour qu'il y ait autant d'hommes que de femmes dans chaque zone.

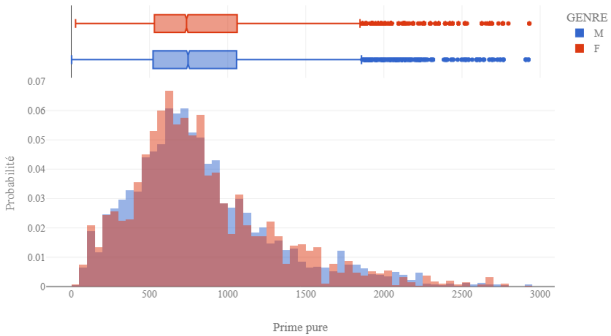


Objectif : égaliser les distributions homme-femme

Mitigation du biais : adaptation fair-SMOTE



Distribution de la prime pure par genre avant fair-SMOTE.



*Distribution de la prime pure par genre **après fair-SMOTE**.*

Le décalage entre les distributions est **réduit**.

Modèles	Référence	Fair-SMOTE
RMSE	171,04	171,61
HGR_KDE	29,71%	28,83%
S/P	99,66%	99,65%

Le biais n'est pas réduit. Les hyperparamètres, les zones et les approches de rééchantillonnage ont été modifié **sans succès**.

Améliorer la répartition des genres ne semble pas améliorer l'équité dans le cas étudié.

Mitigation du biais : exponentiated gradient

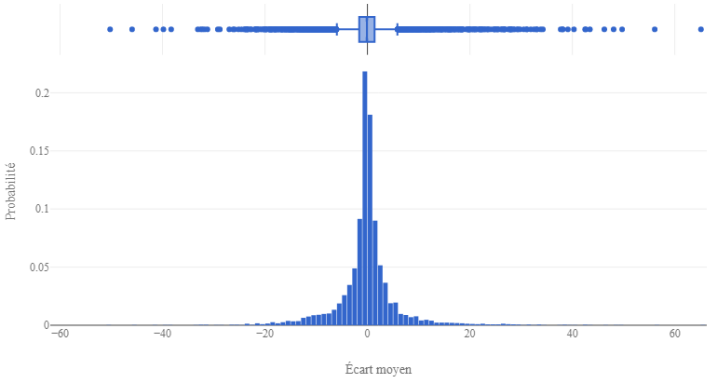
- Implémentation réussie pour une contrainte faible d'équité: **égalité des erreurs par genre**
- Difficulté de généralisation des travaux présents dans la littérature
- Temps restreint pour étudier plus en détails la théorie

Un grand chantier d'étude pour la formulation de méthodes efficientes et adaptées à la tarification.

Modèles	Référence	Exponentiated
RMSE	171,04	171,25
HGR_KDE	29,71%	31,22%
S/P	99,66%	99,65%

Performance du modèle conservée mais sans mitigation du biais.

Mitigation du biais : redistribution équitable



Distribution initiale des écarts

- Biais initiaux mesurés avec l'adaptation du flip-test.
- K plus proches voisins optimisés pour obtenir les distances et les écarts les plus faibles.

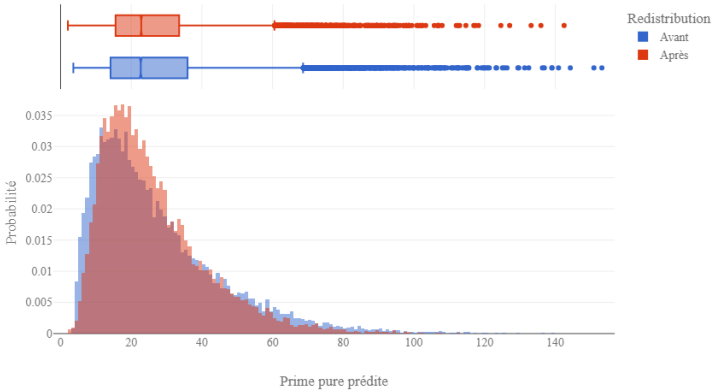
Agrégats	Femme	Homme	Somme
Écart moyen	-0,9€	1,2€	0,3€
Somme des écarts	-25 535€	32 019€	6 484€

Objectif : redistribuer les écarts individuels pour obtenir un état contenant moins de biais.

Arbitrage à faire : une correction trop brutale amplifie le biais et une correction trop lente détruit la cohérence des primes.

Meilleur résultat obtenu en corrigeant 1/5 de l'écart à chaque itération, pour un seuil total d'écart de 1 000€.

Mitigation du biais : redistribution équitable



Primes initiales vs primes après redistribution

✓ Distribution et performances maintenues.
✓ Biais individuels significativement réduits.

\- Attention au biais de modélisation

Avant redistribution

Agrégats	Femme	Homme	Somme
Écart moyen	-0,9€	1,2€	0,3€
Somme des écarts	-25 535€	32 019€	6 484€

S/P 99,5%
RMSE 171,04

Après redistribution

Agrégats	Femme	Homme	Somme
Écart moyen	-0,075€	0,13€	0,3€
Somme des écarts	-850€	2 245€	1 395€

S/P 99,2% soit 1297€ de primes en plus
RMSE 171,66

Récapitulatif

Mesure du biais avant modélisation

Présence du biais dans l'historique

Mesures du biais après modélisation

Présence et parfois amplification du biais quelle que soit l'approche classique utilisée

Mitigation du biais avant modélisation

Mitigation pendant la modélisation

Mitigation après la modélisation

Méthodes de mitigation	Performances (RMSE et S/P)	Niveau de mitigation	Applicabilité	Intérêts
Suppression	✓	✓	✓	Simplicité
Adaptation faire-SMOTE	✓	✗	✓	Problème de représentation
Exponentiated gradient	✓	✗	✗	Erreur de modélisation
Redistribution équitable	✓	✓	✓	Correction individuelle

Conclusion

ETUDE D'UNE PROBLÉMATIQUE D'ACTUALITÉ

- Utilisation de données massives et d'algorithmes complexes.
- De nombreux soucis d'équité dans tous les domaines.

REVUE D'UNE LITTERATURE

- Les éléments présentés sont le fruit d'une revue extensive de la littérature pour tirer le meilleur pour les cas assurantiels.
- Reflexion axée sur la régression et la tarification contrairement à la littérature.

MESURES DU BIAIS

- Vision théorique couplée à une vision métier.
- Présence du biais avant et après modélisation quelle que soit l'approche classique utilisée.
- Impact des interdépendances.

MITIGATION DU BIAIS

- Passage en revue de méthodes cohérentes en tarification.
- Etudes des contraintes et des coûts.