Modélisation du biais dans les recrutements : étude de l'influence d'un biais dans les données d'apprentissage de différentes procédures

Vincent Brault, Alain Lacroux, Philomène Le Gall, Christelle Martin-Lacroux, Angélique Saillet, Shuyu Wang

Modélisation du biais dans les recrutements : étude de l'influence d'un biais dans les données d'apprentissage de différentes procédures

Vincent Brault[1] & Alain Lacroux[2] & Philomène Le Gall[1] & Christelle Martin-Lacroux[3] & Angélique Saillet[1] & Shuyu Wang[1]

[1] *Univ. Grenoble Alpes, CNRS, Grenoble INP*, LJK, 38000 Grenoble, France*
vincent.brault@univ-grenoble-alpes.fr
[2] *Université Paris 1 Panthéon-Sorbonne, PRISM, Paris, France*
[3] *Univ. Grenoble Alpes, Grenoble INP*, CERAG, 38000 Grenoble France*

Résumé. Grâce aux récentes avancées de l'intelligence artificielle, de nouvelles méthodes automatiques pour l'aide à la décision sont mises au service du public (comme la reconnaissance d'images par exemple). Dans la science du recrutement, des solutions automatiques ont également été développées et sont vendues par des start-up. Or, à cause du secret industriel, ces dernières sont évasives sur la technicité des algorithmes proposés. Le but du projet IAB@R, dont est issu ce travail, est de sensibiliser sur la potentielle mauvaise utilisation des méthodes automatiques. Dans cet exposé, nous nous intéressons à l'influence de l'utilisation d'une base biaisée par des algorithmes classiques sur la qualité de prédiction des algorithmes de recrutement.

Mots-clés. Apprentissage statistique, science du recrutement, perceptron multicouche, régression logistique, critère AIC

Abstract. Thanks to recent advances in artificial intelligence, new automatic methods for decision support are being made available to the public (such as image recognition). In the science of recruitment, automatic solutions have also been developed and are sold by start-ups. However, because of industrial secrecy, these start-ups are evasive about the technicality of the proposed algorithms. The goal of the IAB@R project, from which this work stems, is to raise awareness about the potential misuse of automatic methods. In this presentation, we focus on the influence of the use of a database biased by classical algorithms on the prediction quality of recruitment algorithms.

Keywords. Machine learning, recruitment science, multilayer perceptron, logistic regression, AIC criterion

1 Introdution

Les récentes avancées dans le domaine de l'intelligence artificielle furent également l'occasion, notamment, de l'explosion des solutions de recommandations automatiques de candi-

∗. Institute of Engineering Univ. Grenoble Alpes

dats via une recherche sémantique sur les CV [1], des solutions d'analyse de signaux verbaux et non verbaux au cours des entretiens [2], ou encore des outils de prédiction de la capacité des candidats à s'entendre avec une équipe et à travailler efficacement [3]. Toutes ces solutions promettent des recrutements plus efficaces et surtout exempts de biais et de discrimination. Pourtant, si les biais et la discrimination sont des risques inhérents à toute décision de recrutement humain, les particularités qui caractérisent les technologies de l'IA, en termes d'opacité, de complexité, de comportement partiellement autonome ne fournissent pas de garanties pour réduire ces biais, et on peut même penser que l'IA peut engendrer d'autres types de biais (voir Lacroux et Martin-Lacroux (2021)). Par exemple, un algorithme entraîné à prendre des décisions à partir de données sociales biaisées reproduit ces biais, voire les amplifie et crée des discriminations fortes fondées sur des critères comme le sexe, l'âge, l'origine des individus (pour un exemple, voir Besse (2020)).

Dans cet exposé, nous regardons l'influence de l'utilisation d'une base biaisée dans la qualité du choix final d'un CV qui serait exempt de toute discrimination. Pour ce faire, nous commençons par présenter nos schémas de simulations puis les algorithmes étudiés et concluons par une première discussion sur les résultats que nous présenterons le jour de l'exposé.

2 Simulation d'un entretien

Pour la suite, nous simulerons n **entretiens** mettant chacun en concurrence N_{d} **dossiers** $\boldsymbol{d}_{i,j}$ avec $i \in \{1, \dots, n\}$ et $j \in \{1, \dots, N_{\mathrm{d}}\}$. Chaque dossier $\boldsymbol{d}_{i,j}$ est caractérisé par

- K variables **objectives** $(X_{i,j,k})_{1 \leq k \leq K}$,
- K variables **discriminatoires** $(Y_{i,j,k})_{1 \leq k \leq K}$,
- K variables **corrélées** aux variables discriminatoires $(Z_{i,j,k})_{1 \leq k \leq K}$,
- son rang $R_{i,j} \in \{1, \dots, N_{\mathrm{d}}\}$ dans le classement dans l'entretien i,
- sa réussite $W_{i,j} \in \{0, 1\}$ ou pas d'avoir eu le poste i. Comme nous supposons qu'une seule personne est prise au final, $W_{i,j}$ vaut 1 si et seulement si $R_{i,j}$ vaut 1 également.

Pour la suite, nous supposons que les variables $X_{i,j,k}$ sont toutes indépendantes et de même loi gaussienne $\mathcal{N}(0, 1)$. Nous dirons qu'un entretien est **parfait** si le classement n'est effectué que sur la moyenne $\overline{X}_{i,j}$ des variables objectives et nous noterons $R_{i,j}^{(p)}$ le rang du dossier $\boldsymbol{d}_{i,j}$ dans ce classement et $W_{i,j}^{(p)}$ la variable réussite associée. Nous supposons donc qu'il existe une permutation $\sigma_i \in \mathfrak{S}(\{1, \dots, N_{\mathrm{d}}\})$ telle que :

$$\overline{X}_{i,\sigma_i^{-1}(1)} > \overline{X}_{i,\sigma_i^{-1}(2)} > \cdots > \overline{X}_{i,\sigma_i^{-1}(N_{\mathrm{d}})}$$

alors $R_{i,j}^{(p)} = \sigma_i(j)$. Comme les variables $X_{i,j,k}$ sont continues, σ_i est presque sûrement unique pour la mesure de Lebesgue.

1. `https://www.avature.net/fr/`
2. `https://cryfe.swiss/`
3. `https://www.assessfirst.com/fr/solutions/recrutement/`

Par la suite, nous explicitons plusieurs types d'entretiens où les classements sont **biaisés** par les variables discriminatoires ; nous noterons dans ce cas $R_{i,j}^{(b)}$ et $W_{i,j}^{(b)}$ les rangs et réussites associés.

2.1 Entretien biaisé par censure

Dans cette partie, nous simulons le cas où un dossier n'aurait même pas été regardé car la moyenne de ses variables discriminatoires seraient trop élevées par rapport à un seuil $S \in \mathbb{R}$. Ce cas peut survenir lorsqu'un grand nombre de dossiers en lice sont pré-sélectionnés sur la base de critères de niveau pré-établi, comme des seuils minimaux de notes. C'est par exemple le cas dans les processus de recrutement de certaines universités amenées à gérer plusieurs milliers de candidatures. Étant donnée une corrélation $\alpha \in [0, 1[$, nous proposons deux cas :

- Le **cas binaire** où les variables $Y_{i,j,k}$ sont toutes indépendantes, indépendantes des variables $X_{i,j,k}$ et de même loi de Bernoulli $\mathcal{B}(1/2)$. De plus, nous supposons que pour tout triplet (i, j, k), nous avons :
$$Z_{i,j,k} = U_{i,j,k}Y_{i,j,k} + (1 - U_{i,j,k})B_{i,j,k}$$
où les variables $B_{i,j,k}$ sont toutes indépendantes, indépendantes des variables $X_{i,j,k}$ et $Y_{i,j,k}$ et de même loi de Bernoulli $\mathcal{B}(1/2)$ et les variables $U_{i,j,k}$ sont toutes indépendantes, indépendantes des variables $X_{i,j,k}$, $Y_{i,j,k}$ et $B_{i,j,k}$ et de même loi de Bernoulli $\mathcal{B}(\alpha)$.
- Le **cas continu** où les variables $Y_{i,j,k}$ sont toutes indépendantes, indépendantes des variables $X_{i,j,k}$ et de même loi gaussienne $\mathcal{N}(0, 1)$. De plus, nous supposons que pour tout triplet (i, j, k), nous avons :
$$Z_{i,j,k} = \frac{\alpha}{\sqrt{1 - \alpha^2}} Y_{i,j,k} + \varepsilon_{i,j,k}$$
où les variables $\varepsilon_{i,j,k}$ sont toutes indépendantes, indépendantes des variables $X_{i,j,k}$ et $Y_{i,j,k}$ et de même loi gaussienne $\mathcal{N}(0, 1)$.

Dans les deux cas, pour tout triplet (i, j, k), la corrélation entre $Y_{i,j,k}$ et $Z_{i,j,k}$ est α.

Pour chacun des cas, pour chaque entretien i et étant donné un seuil $S \in \mathbb{R}$, nous proposons un classement biaisé de la façon suivante :

1. nous divisons les dossiers en deux groupes : ceux dont la moyenne $\overline{Y}_{i,j}$ est supérieure aux seuils S à savoir l'ensemble $\mathcal{J}_i^{(g)}(S) = \left\{ j \in \{1, \ldots, N_{\mathrm{d}}\} \,\middle|\, \overline{Y}_{i,j} \geq S \right\}$ et ceux dont la moyenne est strictement inférieure $\mathcal{J}_i^{(b)}(S) = \left\{ j \in \{1, \ldots, N_{\mathrm{d}}\} \,\middle|\, \overline{Y}_{i,j} < S \right\}$.
2. nous commençons par ordonner les dossiers de l'ensemble $\mathcal{J}_i^{(g)}(S)$ suivant leurs moyennes $\overline{X}_{i,j}$ donc nous supposons qu'il existe une permutation $\sigma_{i,(g),S} \in \mathfrak{S}\left(\mathcal{J}_i^{(g)}(S)\right)$ telle que :
$$\overline{X}_{i,\sigma_{i,(g),S}^{-1}(1)} > \overline{X}_{i,\sigma_{i,(g),S}^{-1}(2)} > \cdots > \overline{X}_{i,\sigma_{i,(g),S}^{-1}\left(\left|\mathcal{J}_i^{(g)}(S)\right|\right)}$$
puis nous ordonnons les dossiers de l'ensemble $\mathcal{J}_i^{(b)}(S)$ suivant leurs moyennes $\overline{Y}_{i,j}$ donc nous supposons qu'il existe une permutation $\sigma_{i,(b),S} \in \mathfrak{S}\left(\mathcal{J}_i^{(b)}(S)\right)$ telle que :
$$\overline{Y}_{i,\sigma_{i,(b),S}^{-1}(1)} \geq \overline{Y}_{i,\sigma_{i,(b),S}^{-1}(2)} \geq \cdots \geq \overline{Y}_{i,\sigma_{i,(b),S}^{-1}\left(\left|\mathcal{J}_i^{(b)}(S)\right|\right)}.$$

Pour ce deuxième cas, en cas d'égalité (qui peut arriver pour la cas binaire), le choix est laissé au hasard.

3. Au final, nous calculons pour tout dossier $\boldsymbol{d}_{i,j}$ le rang biaisé de la façon suivante :

$$R_{i,j}^{(b)} = \begin{cases} \sigma_{i,(g),S}(j) & \text{si } j \in \mathcal{J}_i^{(g)}(S), \\ \left|\mathcal{J}_i^{(g)}(S)\right| + \sigma_{i,(b),S}(j) & \text{si } j \in \mathcal{J}_i^{(b)}(S). \end{cases}$$

Ainsi, nous supposons que les dossiers *ayant passé le seuil de discrimination* sont classés suivant leurs compétences tandis que les autres dossiers sont classés par *niveau de discrimination.*

2.2 Entretien biaisé par auto-censure

Dans leurs études, Steele et Ambady (2006) montrent que, par exemple, les femmes pouvaient expérimenter un stress lié à l'activation d'un stéréotype de dépréciation genrée quand un test d'aptitude mentionnait que c'était un test mathématique (par rapport à un test identique sans cette mention). Pour mimer ce type de censure, nous reprenons le cas binaire et proposons de prendre un paramètre $\mu > 0$ de **dépréciation** et de calculer pour tout triplet (i, j, k) les variables $\widetilde{X}_{i,j,k}$ dépréciées de la façon suivante :

$$\widetilde{X}_{i,j,k} = X_{i,j,k} - \mu\,(1 - Y_{i,j,k})\,.$$

Dans ce cas, le classement biaisé reprend la technique du classement parfait de la section 2 mais avec les variables $\widetilde{\widetilde{X}}_{i,j}$.

2.3 Anonymisation des dossiers

Enfin, pour les bases d'entraînement, nous fournissons les dossiers complets c'est-à-dire avec les trois variables $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ et les **dossiers anonymisés** avec seulement $(\boldsymbol{X}, \boldsymbol{Z})$. De plus, nous comparons les algorithmes en ayant appris sur les bases parfaites (donc avec les $\boldsymbol{R}^{(p)}$ ou $\boldsymbol{W}^{(p)}$) et les bases biaisées ($\boldsymbol{R}^{(b)}$ ou $\boldsymbol{W}^{(b)}$). Ainsi, les quatre types de données d'apprentissage fournies aux algorithmes peuvent être résumés par le tableau 1.

TABLE 1 – Tableau résumé des quatre configurations pour les bases d'entraînement des algorithmes.

		Classement	
		Parfait	Biaisé
Dossier	Complet	$\left(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{R}^{(p)}, \boldsymbol{W}^{(p)}\right)$	$\left(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{R}^{(b)}, \boldsymbol{W}^{(b)}\right)$
	Anonyme	$\left(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{R}^{(p)}, \boldsymbol{W}^{(p)}\right)$	$\left(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{R}^{(b)}, \boldsymbol{W}^{(b)}\right)$

Pour évaluer la qualité des procédures, nous simulons m entretiens $i \in \{n+1, \ldots, n+m\}$ de N_{d} dossiers également et nous fournissons les valeurs $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ et $(\boldsymbol{X}, \boldsymbol{Z})$ suivant si les dossiers sont complets ou anonymes et estimons la qualité des résultats sur les classements parfaits.

3 Algorithmes étudiés

Plusieurs méthodes ont été explorées mais dans cet exposé, nous présenterons les trois procédures suivantes :

- la régression logistique (voir Berkson (1944)),
- la régression logistique avec sélection de variables par le critère *AIC* (*Akaike Information Criterion* ; voir Akaike (1973)),
- le perceptron multicouche (voir Rosenblatt (1957)).

Ce choix a été fait pour deux raisons :

1. comme nous le verrons, lorsque les classements parfaits sont utilisés pour l'apprentissage, ces méthodes donnent des taux de bons classements supérieurs à 80% contrairement, par exemple, à l'algorithme *CART* (*Classification And Regression Trees*) de Breiman et al. (1984).
2. une complexité croissante entre les différentes méthodes afin de voir comment cette dernière peut intervenir dans la sensibilité aux données d'apprentissage.

Dans le suite de cette section, nous rappelons quelques notions pour chacune des méthodes. Pour l'apprentissage, nous supposons que les méthodes n'ont accès qu'au fait que les dossiers aient eu le poste ou pas (et pas leur classement).

3.1 Régression logistique

La **régression logistique** suppose que la variable W est issue d'une loi de Bernoulli de probabilité dépendante des variable $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$. En particulier, elle suppose qu'il existe un vecteur $\boldsymbol{\beta} \in \mathbb{R}^{3K+1}$ telle que :

$$\begin{aligned}\mathbb{P}\left(W = 1 \,|\boldsymbol{d}; \boldsymbol{\beta}\right) = \mathbb{P}\left(W = 1 \,|X_1, \ldots, X_K, Y_1, \ldots, Y_K, Z_1, \ldots, Z_K; \boldsymbol{\beta}\right) \\ = \Phi\left[\beta_0 + \sum_{k=1}^{K} \left(\beta_k X_k + \beta_{K+k} Y_k + \beta_{2K+k} Z_k\right)\right]\end{aligned}$$

où Φ est la fonction logistique définie pour tout $x \in \mathbb{R}$ par :

$$\Phi\left(x\right) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}.$$

Ainsi, étant donnée une estimation $\widehat{\boldsymbol{\beta}}$ du vecteur $\boldsymbol{\beta}$ et de nouveaux dossiers à un entretien $i \in \{n+1, \ldots, n+m\}$, nous pouvons calculer pour chaque dossier $\boldsymbol{d}_{i,j}$ sa probabilité d'avoir

l'entretien et nous classons les dossiers de la plus grande probabilité à la plus petite. Nous supposons qu'il existe une permutation $\sigma_i \in \mathfrak{S}(\{1,\dots,N_{\mathrm{d}}\})$ telle que :

$$\mathbb{P}\left(W=1\left|\boldsymbol{d}_{i,\sigma_i^{-1}(1)};\widehat{\boldsymbol{\beta}}\right.\right) > \mathbb{P}\left(W=1\left|\boldsymbol{d}_{i,\sigma_i^{-1}(2)};\widehat{\boldsymbol{\beta}}\right.\right) > \cdots > \mathbb{P}\left(W=1\left|\boldsymbol{d}_{i,\sigma_i^{-1}(N_{\mathrm{d}})};\widehat{\boldsymbol{\beta}}\right.\right)$$

et on définit $\widehat{R}_{i,j} = \sigma_i(j)$.

Pour ce faire, nous utilisons la fonction `glm` du package `stats` du logiciel R Core Team (2022).

3.2 Régression logistique et critère *AIC*

Le fait de prendre toutes les variables comme dans la section 3.1 peut entraîner du surapprentissage et il est donc conseillé d'utiliser une méthode de sélection de variables. Comme nous sommes dans un problème de prédiction, nous avons choisi d'utiliser le critère *AIC* qui sélectionne les variables maximisant la logvraisemblance diminuée du nombre de paramètres considérés comme non nuls :

$$\widehat{\boldsymbol{\beta}}_{\mathrm{AIC}} \in \underset{\boldsymbol{\beta}\in\mathbb{R}^{3K+1}}{\operatorname{argmax}} \left\{ \mathcal{L}(\boldsymbol{\beta}) - \sum_{k=0}^{3K+1} \mathbb{1}_{\{\beta_k\neq 0\}} \right\}$$

où $\mathcal{L}(\boldsymbol{\beta})$ est la logvraisemblance et $\mathbb{1}$ est la fonction indicatrice.

Comme cela représente 2^{3K+1} modèles possibles à comparer, il devient très vite chronophage de tester tous les modèles (65536 pour $K=5$) voir impossible de réaliser en moins d'une année pour le supercalculateur *Summit*[4] dès que K est plus grand que 27. Pour contourner ce problème de temps, nous pouvons utiliser

- la méthode *forward* qui part du vecteur nul et autorise progressivement des coordonnées du vecteur $\boldsymbol{\beta}$ à être non nulles en sélectionnant à chaque fois celle qui augmente le plus le critère *AIC* jusqu'à la stabilisation de ce dernier ;
- la méthode *backward* qui part du vecteur complet et met progressivement à zéro des coordonnées du vecteur $\boldsymbol{\beta}$ en sélectionnant à chaque fois celle qui augmente le plus le critère *AIC* jusqu'à la stabilisation de ce dernier.

Pour la procédure testée, nous utilisons la combinaison des deux implémentée dans la fonction `stepAIC` du package `MASS` de `R` (voir Venables et Ripley (2002)).

3.3 Perceptron multicouche

La dernière procédure utilisée est celle du perceptron multicouche. Le principe d'un perceptron à une couche ou perceptron simple ou neurone consiste à supposer qu'il existe une fonction g et un paramètre $\boldsymbol{\beta}\in\mathbb{R}^{3K+1}$ tels que

$$W = g\left[\beta_0 + \sum_{k=1}^{K}\left(\beta_k X_k + \beta_{K+k} Y_k + \beta_{2K+k} Z_k\right)\right].$$

4. 148,6 pétaflops ; en admettant qu'il traite un modèle par opération

En particulier, si $g = \Phi$ la fonction logistique, nous retrouvons le modèle logistique de la section 3.1 (voir une schématisation sur la partie gauche de la figure 1).

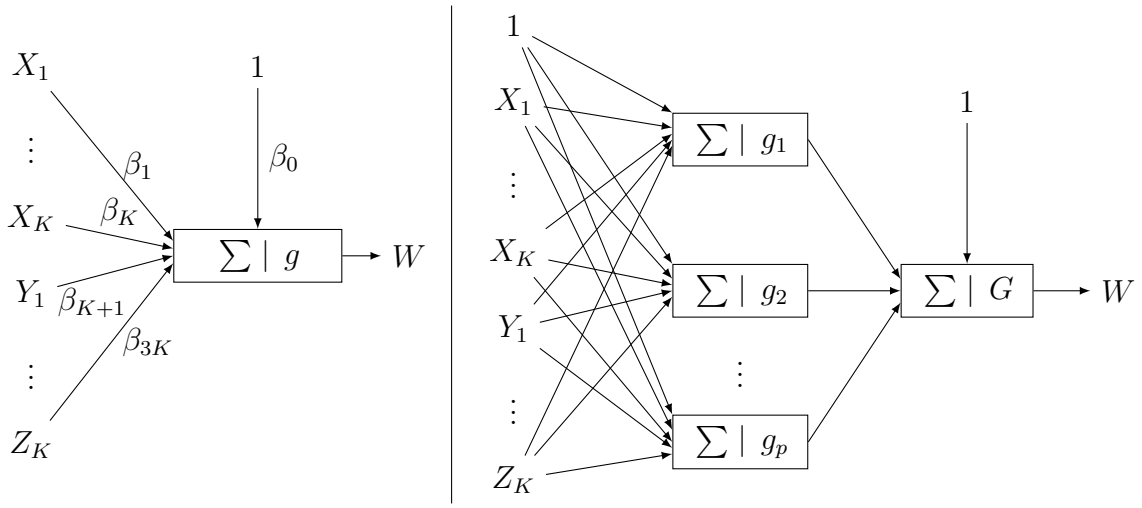

FIGURE 1 – Représentation schématique d'un perceptron à une couche (à gauche) et d'un perceptron multicouche avec une couche cachée contenant p neurones (à droite).

Un **perceptron multicouche** est la combinaison de plusieurs perceptrons simples : chaque couche contient un certain nombre de perceptrons simples prenant en entrée soit les données initiales, soit les transformations provenant des couches précédentes.

Dans le schéma de droite de la figure 1, nous avons représenté le cas où il n'y aurait qu'une seule couche cachée contenant p perceptrons simples. Si chaque fonction g_k est la fonction valant 0 ou l'entrée k (X, Y ou Z) et $G = \Phi$ la fonction logistique alors nous retrouvons l'objectif de la procédure de la section 3.2.

Vu le nombre de possibilités, nous n'étudions que le cas d'un perceptron avec une couche cachée ne contenant que 2 neurones et optimisons le tout à l'aide de la fonction `nnet` du package du même nom de `R` (voir Venables et Ripley (2002)).

4 Plans de simulations

Nous avons choisi de simuler $n = 1000$ entretiens pour la base d'entraînement et $m = n/10$ pour la base de tests composés chacun de $N_{\mathrm{d}} = 5$ dossiers. De plus, nous prenons $\alpha \in \{0.2, 0.5, 0.8\}$. Dans les études en management (voir la méta-analyse de Paterson et al. (2016)), 0.2 est considéré comme une corrélation correspondant à une taille d'effet moyenne, 0.5 comme une corrélation correspondant à un effet important et 0.8 comme une corrélation correspondant à un effet très important, quasiment jamais rencontré.

Afin d'avoir la même probabilité que deux dossiers soient discriminés dans les plans quantitatifs et qualitatifs, nous avons fixé pour le second les seuils $S \in \{0, \ldots, 4/5\}$ et, pour le premier, les seuils dans $S \in \{q_{\mathbb{P}(C\leq 0)}, \ldots, q_{\mathbb{P}(C\leq 4)}\}$ où q est le quantile de la loi gaussienne centrée de variance $1/N_{\mathrm{d}}$ et C suit une loi binomiale de paramètres $\mathcal{B}in\,(N_{\mathrm{d}}; 1/2)$; ainsi nous

obtenons des probabilités de rejets valant environ $\{3\%, 19\%, 50\%, 81\%, 97\%\}$ dans les deux cas. Pour les valeurs de μ, nous avons choisi $\mu \in \{0.4, 0.8, \ldots, 2\}$.

De plus, chaque dossier est composé de $K = 5$ variables de chaque (soit 15 en tout). Afin de limiter la part d'aléatoire, nous simulons à chaque fois les valeurs intrinsèques au scénario (par exemple, $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\varepsilon})$ pour les plans quantitatifs et $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{B})$ pour les plans qualitatifs).

Enfin, pour chaque configuration, nous simulons 1000 jeux de données et donnons aux algorithmes uniquement les dossiers et s'ils ont eu l'entretien pour la base d'entrainement (sans dire quels dossiers ont été mis en concurrence). Pour chaque entretien de la base de test, nous regardons quel dossier a été sélectionné et comparons avec le dossier choisi par le classement *parfait*.

Dans cet exposé, nous détaillerons les résultats obtenus. Sur la figure 2, nous avons représenté les résultats du perceptron multicouche pour le scénario quantitatif avec seuil de discrimination. Plus les points sont à droite, meilleure est la classification avec la base d'entraînement avec le classement parfait et plus les points sont hauts, meilleure est la classification avec les bases biaisées.

Comme les points sont globalement en-dessous de la droite $y = x$, il y a une dégradation dans l'estimation si l'apprentissage se fait sur une base biaisée.

La séparation des ellipses permet de mesurer l'intérêt de la mise en place des dossiers anonymes. Dans le cas d'une faible corrélation, les ellipses sont séparées et nous observons que l'anonymisation permet d'améliorer les résultats. Par contre, cette intérêt diminue pour une corrélation forte ($\alpha = 0.5$) et s'il existait des corrélations très fortes (0.8) alors l'intérêt serait inexistant.

5 Discussions

Dans cet exposé, nous comparerons les résultats obtenus par les différentes procédures sur les différents scénarios et discuterons des implications sur la circulation des algorithmes proposés actuellement.

6 Remerciements

Ce travail est soutenu par l'Agence nationale de la recherche française dans le cadre du programme "Investissements d'avenir" (ANR-15-IDEX-02). Tous les calculs présentés dans ce document ont été effectués à l'aide de l'infrastructure GRICAD (`https://gricad.univ-grenoble-alpes.fr`), qui est soutenue par les communautés de recherche de Grenoble.

Références

H. Akaike. Information theory and an extension of the maximum likelihood principle. Dans Proceedings, 2nd Internat. Symp. on Information Theory, page 267–281, 1973.

J. Berkson. Application of the logistic function to bio-assay. Journal of the American statistical association, 39(227) :357–365, 1944.

P. Besse. Détecter, évaluer les risques des impacts discriminatoires des algorithmes d'ia. 2020.

L. Breiman, J. Friedman, R. Olshen, et C. Stone. Classification and regression trees. wadsworth int. Group, 37(15) :237–251, 1984.

A. Lacroux et C. Martin-Lacroux. L'intelligence artificielle au service de la lutte contre les discriminations dans le recrutement : nouvelles promesses et nouveaux risques. Management Avenir, (2) :121–142, 2021.

T. A. Paterson, P. Harms, P. Steel, et M. Credé. An assessment of the magnitude of effect sizes : Evidence from 30 years of meta-analysis in management. Journal of Leadership & Organizational Studies, 23(1) :66–81, 2016.

R Core Team. R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL `https://www.R-project.org/`.

F. Rosenblatt. The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory, 1957.

J. R. Steele et N. Ambady. "math is hard !" the effect of gender priming on women's attitudes. Journal of Experimental Social Psychology, 42(4) :428–436, 2006.

W. N. Venables et B. D. Ripley. Modern Applied Statistics with S. Springer, New York, fourth edition, 2002. URL `https://www.stats.ox.ac.uk/pub/MASS4/`. ISBN 0-387-95457-0.

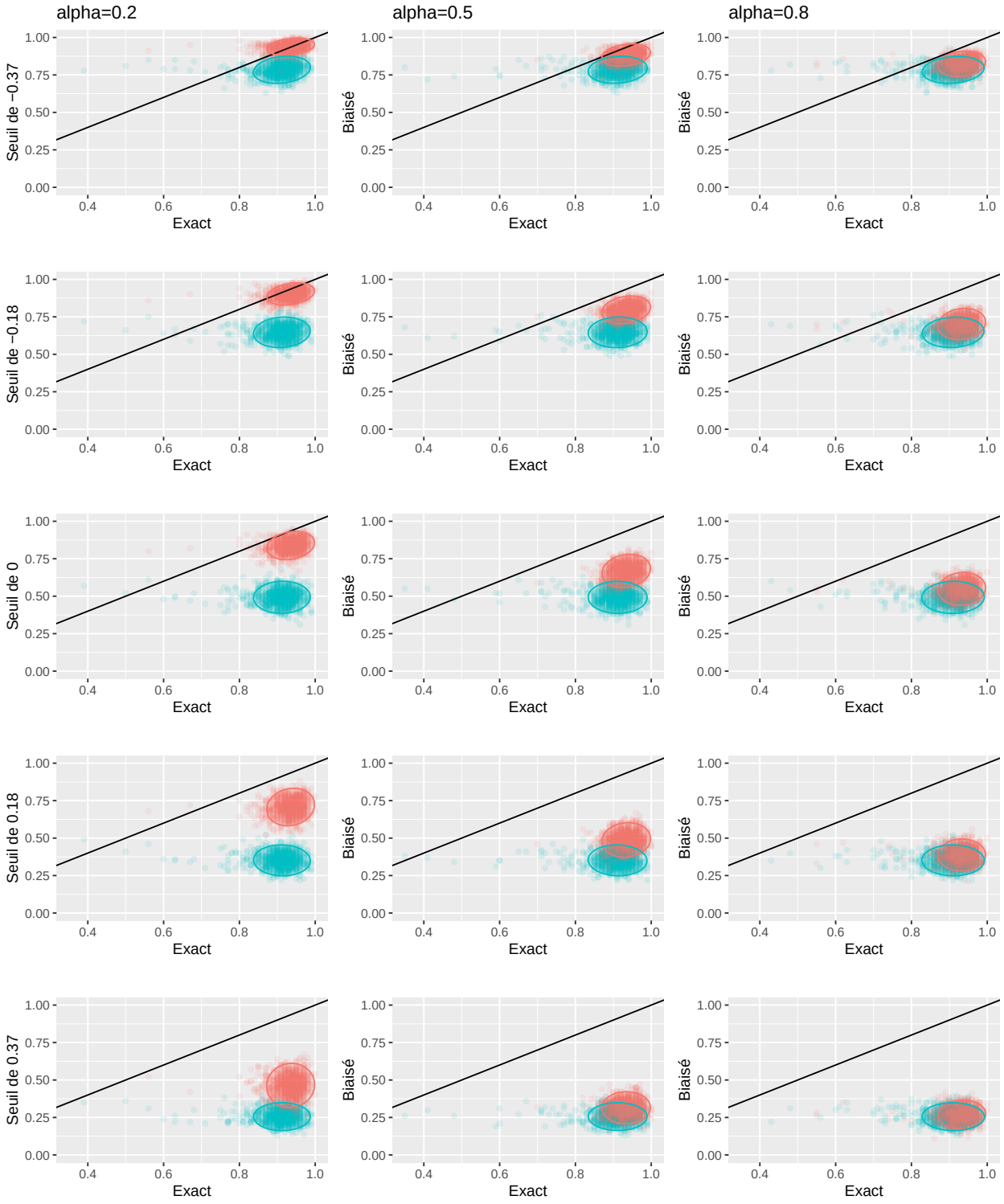

FIGURE 2 – Représentation des résultats issus du perceptron multicouche où chaque point a pour abscisse le taux moyen de bonnes classifications si l'algorithme avait le classement parfait et en ordonnée le classement biaisé suivant le seuil de discrimination (lignes), la corrélation α (colonnes) et le type de dossier (en turquoise pour les dossiers complets et saumon pour les dossiers anonymisés. La droite $y = x$ est ajoutée en noir et des ellipses à 95% ont été ajoutées.