UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RÉGRESSION PAR DISCONTINUITÉ SUR UNE VARIABLE RÉPONSE BINAIRE : L'EFFET DE L'ÂGE RELATIF D'UN ENFANT SUR LA PROBABILITÉ D'ÊTRE DIAGNOSTIQUÉ DU TROUBLE DÉFICITAIRE DE L'ATTENTION AVEC OU SANS HYPERACTIVITÉ (TDAH)

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

FRANCELYNE JEAN-BAPTISTE

MARS 2020

REMERCIEMENTS

Merci à ma directrice Geneviève Lefebvre.

DÉDICACE

Je dédie ce mémoire à ma mère, ma tante et mes soeurs. Ces femmes fortes qui m'ont montré que la détermination, la concentration et la passion étaient les clés de la réussite. Ces femmes qui m'ont encouragé, lors de nombreuses nuits blanches, à persévérer et à ne pas lâcher. Ces femmes qui, de leurs nombreuses prières, m'ont montré, encore aujourd'hui, ce que la foi peut surmonter.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

RÉSUMÉ

Dans le cadre de ce mémoire, nous étudions, à l'aide de simulations, la régression par discontinuité pour une variable réponse binaire. Nous débutons par la définition du modèle et de ses deux composantes ("sharp" et "fuzzy"), pour ensuite le définir dans un contexte contrefactuel. Puis nous décrivons les hypothèses dans le cas "sharp" et présentons l'estimation, pour une réponse continue, de l'effet moyen du traitement à l'aide de deux procédures : l'approche par régression linéaire fréquentiste et bayésienne. Par la suite, nous nous concentrons sur le cas d'une réponse binaire dans le contexte de l'étude d'association entre l'âge d'entrée des enfants à l'école et la probabilité d'être diagnostiqué du trouble déficitaire de l'attention avec ou sans hyperactivité (TDAH). Pour ce faire, nous évaluons la capacité du modèle à estimer adéquatement l'effet du traitement pour différentes tailles d'échantillon et de largeurs de bande. Puis nous explorons l'impact de l'ajout d'une variable explicative et de confusion sur l'ajustement du modèle. Finalement, nous utilisons la simulation bayésienne pour analyser les conséquences du choix de la loi a priori des coefficients de régression sur l'estimation du paramètre d'intérêt et comparons cette approche avec celle fréquentiste. Nous terminons par un retour sur nos résultats et une discussion ouverte sur d'autres problématiques en lien avec la régression par discontinuité dans d'autres secteurs de recherche.

Mots clés : régression par discontinuité, régression par discontinuité "sharp", réponse binaire, simulation bayésienne, TDAH, rstan, MCMC

INTRODUCTION

Le modèle de régression par discontinuité (RD) est une méthode quasi-expérimentale introduite dans les années 60 par (Thistlethwaite et Campbell, 1960). Un de ses avantages est qu'il peut être appliqué dans tout contexte où une intervention ou un traitement est administré selon une loi prédéterminée liée à une variable continue dite variable d'affectation. Depuis les années 90, la RD a été fréquemment adoptée dans le domaine de l'économie (Imbens et Lemieux, 2008) et, plus récemment, en épidémiologie et en sciences médicales ((Geneletti *et al.*, 2016), (O'Keeffe et Baio, 2016), (Geneletti *et al.*, 2015), (Bor *et al.*, 2014)). La caractéristique principale de la RD est que les individus situés juste en dessous du point de discontinuité sont comparables avec ceux situés juste au dessus. En se concentrant autour de ce point, nous nous retrouvons avec une situation analogue à un essai randomisé contrôlé sans les coûts, ni les complications qui s'y rattachent.

Dans ce mémoire, nous nous intéressons à l'application du modèle de régression par discontinuité pour une réponse binaire. La plus grande partie de la littérature sur la RD traite de cas d'une variable de réponse continue. Entre autres, (Bor *et al.*, 2014) nous présente le modèle RD et les hypothèses requises pour son implémentation dans un contexte d'inférence causale, se référant plus particulièrement aux cas d'analyses épidémiologiques (modèle non linéaire et de survie). À notre connaissance, (Geneletti *et al.*, 2016) est le seul article qui détaille l'approche bayésienne dans le cas d'une réponse dichotomique. Le but de ce mémoire est de fournir des balises supplémentaires pour comprendre et implémenter le modèle de régression par discontinuité pour une réponse binaire qui, dans le contexte épidémiologique et des soins de santé, est d'un grand intérêt.

Tout au long de ce mémoire, nous ferons référence à une mise en contexte en particulier, soit l'étude de l'association entre la probabilité d'être diagnostiqué avec un trouble déficitaire de l'attention avec ou sans hyperactivité (TDAH) et l'âge d'entrée des enfants à l'école. Dû à la date éligible d'entrée à l'école, les enfants dans une même classe peuvent avoir presque un an de différence d'âge, tel que ceux nés juste avant la date limite seront plus jeunes et peut-être moins matures que leurs camarades de classe nés après cette date (L Morrow *et al.*, 2012). Prenons l'exemple de deux enfants admis à la maternelle dans la même classe : l'enfant A est âgé de 4 ans et 11 mois et l'enfant B est âgé de 5 ans et 10 mois à la date éligible d'entrée. Ainsi, l'élève A plus jeune pourrait être diagnostiqué TDAH à cause de son comportement jugé immature relativement à l'élève B. Par contre, si l'enfant A était né quelques jours plus tôt, il serait admis l'année suivante et alors sa conduite serait considérée comme étant normale par rapport aux autres élèves plus vieux de la classe ; il n'y aurait pas de questionnement quant à la possiblilté d'être atteint de TDAH. Ainsi, ce sont ces quelques jours d'écart qui font toute la différence pour un enfant entre la possibilité d'être diagnostiqué ou non TDAH.

Plusieurs études, au niveau international, ont rapporté une association entre ces deux facteurs : Canada (L Morrow *et al.*, 2012), États-Unis ((E Elder, 2010), (N Evans *et al.*, 2010)), Suède (Halldner *et al.*, 2014) et Allemagne (Krabbe *et al.*, 2014). Le Québec n'échappe pas à ce phénomène de surdiagnostic où, de 2006-2014, le nombre d'unités de médicaments spécifiques au TDAH est passé de 23 à 65 millions (secteur privé et public), tandis que les coûts associés sont passés de 34 à 160 millions $ pour la Régie de l'assurance maladie du Québec (RAMQ) et de 24 à 116 millions $ chez les assureurs privés (Turgeon, 2017). Ce mémoire se veut un outil qui pourrait servir, entre autres, à analyser ce type de phénomène où la réponse est binaire.

Dans le premier chapitre, nous détaillons les propriétés du modèle de régression

par discontinuité dans un contexte de général. Nous débutons par faire la distinction entre les deux types de modèle RD (section 1.1.1), tout en introduisant la notation causale contrefactuelle liée au modèle (sections 1.1.2-1.1.3) et le concept de confusion (section 1.1.4). Nous poursuivons en mettant en place les balises pour l'application du modèle par discontinuité "sharp". Pour ce faire, nous expliquons les hypothèses nécessaires à l'utilisation du modèle et donnons une première définition de l'effet moyen du traitement (EMT) (section 1.2). Par la suite, nous présentons deux procédures pour estimer l'EMT dans le cas d'une réponse continue, soit la régression linéaire simple fréquentiste (section 1.3.1) et sa contrepartie bayésienne (section 1.3.2). Par la suite, nous introduisons deux types de mesures de l'EMT dans le cas d'une réponse binaire : le risque relatif causal et le rapport de cotes causal (section 1.4).

Au deuxième chapitre, nous décrivons l'étude de simulation principale en débutant par détailler l'algorithme pour générer une base de données d'origine (section 2.1), suivie de l'étude de l'impact de la taille d'échantillon (section 2.2.1) et de la valeur d'un paramètre de réglage (largeur de bande) sur l'estimation de l'EMT (section 2.2.2) ; ce dernier étant un élément nécessaire au roulement de l'algorithme. Ensuite, nous ajoutons une variable explicative (section 2.3) puis confondante (section 2.4) pour analyser leur impact sur l'ETM estimé en fonction de la taille d'échantillon et de la largeur de bande.

Notre dernier chapitre porte sur l'inférence bayésienne du modèle RD et nous commençons par expliquer les objectifs liés à son utilisation (section 3.1). Nous introduisons, entre autres, quelques notions de base (section 3.1.1) sur les techniques d'échantillonnage Monte-Carlo par chaînes de Markov (MCMC) et justifions nos décisons quant aux lois a priori utilisées (section 3.2). Nous poursuivons par une introduction au Monte-Carlo Haniltonien (section 3.3.1) et appliquons un diagnostic de convergence pour une implémentation optimale du modèle bayésien

(3.3.2). Finalement, nous terminons par la présentation des résultats et l'analyse prédictive de la loi a posteriori (section 3.4).

Nous concluons notre mémoire par un bilan des résultats obtenus et proposons différentes extensions du modèle de régression par discontinuité qui seraient intéressantes à examiner au-delà de ce mémoire.

CHAPITRE I

MODÈLES DE RÉGRESSION PAR DISCONTINUITÉ

Nous traitons dans ce chapitre de la définition du modèle régression par discontinuité (RD) et de ses deux composantes ("sharp" et "fuzzy"). Par la suite, nous introduisons la notation contrefactuelle et terminons par une définition de la confusion dans un contexte RD. Nous poursuivons en traitant du cas général de variable de réponse continue, pour finalement se concentrer sur le cas de réponse binaire.

1.1 Régression par discontinuité et inférence causale

1.1.1 Introduction au modèle RD

Comme mentionné précedemment, la régression par discontinuité (RD) est une méthode quasi-expérimentale dont le but est d'effectuer une inférence causale par l'estimation de l'effet d'un traitement ou d'une intervention sur une variable réponse d'intérêt. Ce modèle se voit appliqué dans des situations où les individus reçoivent le traitement selon leur position de part et d'autre d'un seuil fixé selon une variable d'affectation continue. Par conséquent, l'affectation du traitement est déterminée soit totalement ou partiellement par le respect de cette règle de décision, tel que toute discontinuité dans la loi conditionnelle de la variable réponse est l'indication d'un effet causal de traitement (Imbens et Lemieux, 2008).

Il existe deux modèles RD principaux : le modèle RD "sharp" (RDS) et le modèle RD "fuzzy" (RDF), qui se distinguent par le respect de la règle de décision. Le RDS est utilisé lorsque la règle de décision est parfaitement respectée, tel que tous les individus situés au-dessus du seuil reçoivent le traitement et ceux en-dessous ne le reçoivent pas. Dans le cas du RDF, la règle de décision est partiellement respectée : certains reçoivent le traitement contrairement à ce qu'indique la règle de décision et d'autres ne le reçoivent pas même si la règle indique qu'ils devraient en bénéficier.

Prenons par exemple le cas épidémiologique de la prescription de statines, un médicament utilisé dans la prévention des maladies cardiovasculaires. Le National Institute for Health and Care Excellence (NICE) recommande de prescrire des statines aux patients n'ayant pas eu d'incidents cardiovasculaires si le risque de développer une maladie cardiovasculaire, au cours des dix prochaines années, excède 20% (O'Keeffe et Baio, 2016) ; (Geneletti *et al.*, 2015). Un individu recevra le traitement (prescription de statines) si la variable continue d'affection (taux de risque de maladies cardiovasculaires) est plus grande ou égale à un seuil fixe (20%) ; dans le cas contraire, il ne sera pas sujet au traitement. Cet exemple fait directement référence au cas d'un modèle RDS.

Le choix du seuil est prédéfini et s'appuie habituellement sur des présuppositions a priori résultant d'une étude déjà publiée ou de recherches approfondies dans la littérature. Rappelons que l'une des hypothèses indispensables à la RD est l'idée d'échangeabilité, soit que les individus situés juste au-dessus du seuil et ceux juste au-dessous ont des caractéristiques similaires. Il nous est ainsi possible de comparer les résultats entre ces deux groupes pour avoir une évaluation appropriée de l'effet causal de l'intervention ou du traitement sur la variable réponse d'intérêt (O'Keeffe et Baio, 2016).

Notons que pour illustrer le modèle RDS, nous avons utilisé un exemple connu dans la littérature bayésienne. Dans celui-ci, la variable réponse est continue, alors que nous allons considérer, au cours de ce mémoire, un cas de variable réponse binaire. Les concepts d'application demeurent toutefois similaires.

1.1.2 Modèle contrefactuel

Dans le cas d'une intervention ou d'un traitement binaire, nous conceptualisons, pour chaque sujet i, deux états possibles. Soit Y_i une variable réponse d'intérêt associée au sujet i, alors Y_{i1} est la variable réponse pour le sujet i lorsque celui-ci est exposé au traitement et Y_{i0} est la variable réponse pour ce sujet lorsqu'il n'est pas exposé au traitement.

Nous introduisons la variable indicatrice de traitement T_i égale à 1 si le sujet i est exposé au traitement et égale à 0 si le sujet i n'est pas exposé. Ainsi, nous pouvons redéfinir la variable continue d'intérêt :

$$Y_i = \begin{cases} Y_{i0} & \text{si } T_i = 0 \\ Y_{i1} & \text{si } T_i = 1 \end{cases} = T_i \cdot Y_{i1} + (1 - T_i) \cdot Y_{i0}. \tag{1}$$

Cette définition est possible grâce à l'hypothèse de consistance ("consistency") qui stipule que la variable réponse contrefactuelle d'un individu exposé au traitement (Y_{i1}) est égale à la variable réponse observée de celui-ci (Y_i) s'il a réellement reçu le traitement auquel il a été exposé (Oldenburg *et al.*, 2016). La même logique s'applique pour un individu non exposé au traitement.

Dans le cas d'une variable réponse continue, nous définissons typiquement l'effet causal individuel comme étant le contraste suivant entre Y_{i1} et Y_{i0}, soit $Y_{i1} - Y_{i0}$. Or, il est impossible d'observer Y_{i1} et Y_{i0} simultanément pour un seul sujet,

par conséquent, la variable réponse observée (Y_i), dans l'équation (1), détient seulement une partie de l'information nous permettant de définir l'effet causal individuel (Morgan et Winship, 2007). En effet, pour un individu traité ($T_i = 1$), nous ne pouvons observer la variable contrefactuelle Y_{i0} et pour un individu non traité ($T_i = 0$), nous ne pouvons observer la variable contrefactuelle Y_{i1}.

Face à cette situation, l'alternative est d'utiliser l'effet moyen du traitement (EMT) dans la population (i.e. sur l'ensemble des individus). Dans le contexte de la régression par discontinuité, il est utile de définir celui-ci conditionnellement à la variable d'affectation X_i :

$$\begin{aligned} EMT_x &= E[Y_{i1} - Y_{i0}|X_i = x] \\ &= E[Y_{i1}|X_i = x] - E[Y_{i0}|X_i = x]. \end{aligned}$$

L'EMT mesure donc la différence en moyenne entre la réponse contrefactuelle des sujets qui reçoivent le traitement et la réponse contrefactuelle de ceux qui ne le reçoivent pas pour une valeur de x donnée. Mentionnons que l'EMT n'est pas équivalent à $E[Y_i|T_i = 1, X_i = x] - E[Y_i|T_i = 0, X_i = x]$ qui est le contraste de l'espérance de la réponse observée (Y_i) conditionnelle à la variable de traitement (T_i) pour une valeur de x donnée.

1.1.3 Modélisation

Soient X_i la variable continue d'affectation, S_i la variable indicatrice du seuil, x_0 le seuil de décision fixe et T_i la variable indicatrice de traitement tels que :

$$S_i = \begin{cases} 1 & \text{si } X_i \geq x_0 \\ 0 & \text{si } X_i < x_0 \end{cases}$$

$$T_i = \begin{cases} 1 & \text{si le } i\text{ème sujet reçoit le traitement} \\ 0 & \text{si le } i\text{ème sujet ne reçoit pas le traitement .} \end{cases}$$

Pour chaque individu i, nous observons (Y_i,X_i,S_i,T_i,C_i), où C_i représente l'ensemble des variables confondantes. En référence à l'exemple introductif, le sujet recevra une prescription de statines (T_i) si son taux de risque (X_i) est plus élevé que le seuil préétabli de 20% (x_0).

1.1.4 Confusion

Avant de poursuivre notre description du modèle RDS, nous introduisons le concept de confusion qui désigne une situation où la présence d'une troisième variable vient fausser la relation observée entre la variable explicative d'intérêt et la réponse (Joseph, 2019). Ainsi, C_i est une variable confondante et donc une cause commune de la variable explicative d'intérêt, T_i, et la réponse Y_i. La figure 1.1 illustre ce phénomène dans un cas de généralité.

Rappelons que, dans notre contexte de RDS, C_i ne peut avoir de lien direct à T_i, car cette dernière est une fonction déterministe de la variable d'affectation X_i. Ainsi, il n'existe pas de confusion survenant directement de la variable confondante tel qu'illustré à la figure 1.1. Nous pouvons toutefois introduire de la confusion, à travers un lien entre C_i et X_i, tel qu'illustré à la figure 1.2. Il s'en suit qu'il existe une association non causale entre T_i et Y_i due à la présence du trajet porte-arrière ("backdoor path") $T_i \leftarrow X_i \leftarrow C_i \rightarrow Y_i$ (Hernán et Robins, 2019).

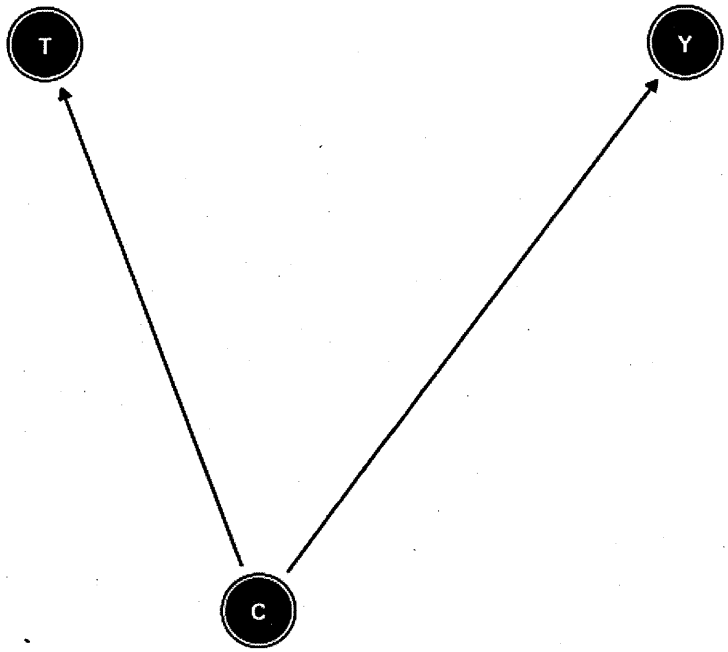

Figure 1.1: Graphique orienté acyclique causal ("DAG") représentant les associations avec la variable de confusion

Pour le reste du mémoire, jusqu'à indication contraire, nous ferons l'hypothèse qu'il n'y a pas de confusion dans le contexte RDS que nous développerons.

1.2 Modèle RD "sharp" et ses hypothèses

Tel que mentionné précédemment, on fait référence au modèle RDS si la règle de décision est strictement respectée, de façon que les sujets reçoivent le traitement si la variable d'affectation est plus grande ou égale au seuil fixe, $X_i \geq x_0$, tandis que les sujets avec une variable d'affectation plus petite, $X_i < x_0$, ne le reçoivent pas. Le modèle RDS présuppose donc que la fonction indicatrice du traitement est une fonction déterministe de la variable d'affectation :

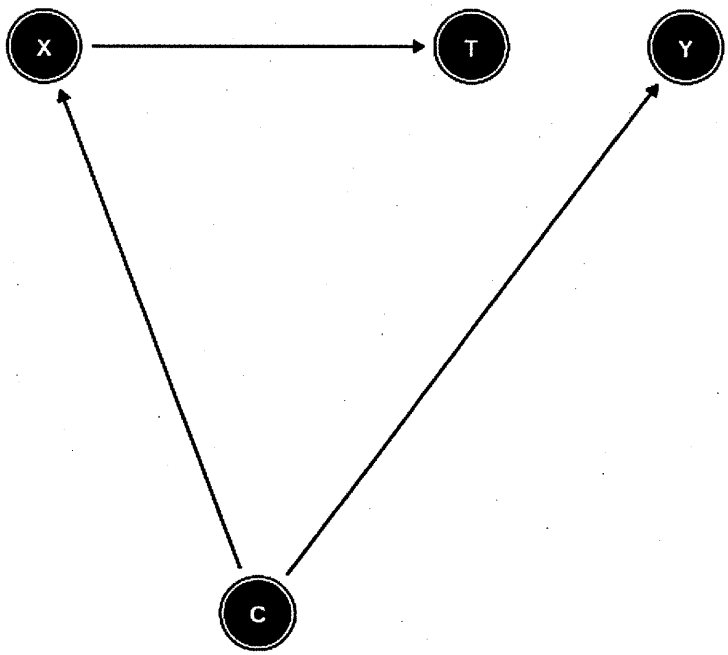

Figure 1.2: Graphique orienté acyclique causal ("DAG") représentant les associations avec la variable de confusion dans un modèle RDS

$$T_i = \begin{cases} 1 & \text{si } X_i \geq x_0 \\ 0 & \text{si } X_i < x_0. \end{cases}$$

Il s'ensuit que la variable indicatrice de seuil, S_i, et la variable indicatrice du traitement, T_i, jouent le même rôle pour l'estimation de l'effet moyen du traitement à l'aide du modèle RDS.

Les figures 1.3 et 1.4 nous permettent de résumer le concept général du modèle RDS. Dans la figure 1.3, la variable d'affectation, X_i, est sur l'axe des abscisses et la probabilité de recevoir le traitement, $P(T_i = 1|X_i = x)$, sur l'axe des ordonnées. Le point de discontinuité est au point $x_0 = 20$. Nous observons, dans la

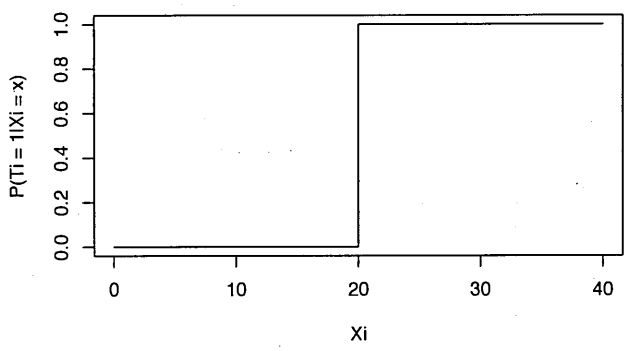

Figure 1.3: Probabilité d'affectation du traitement en %

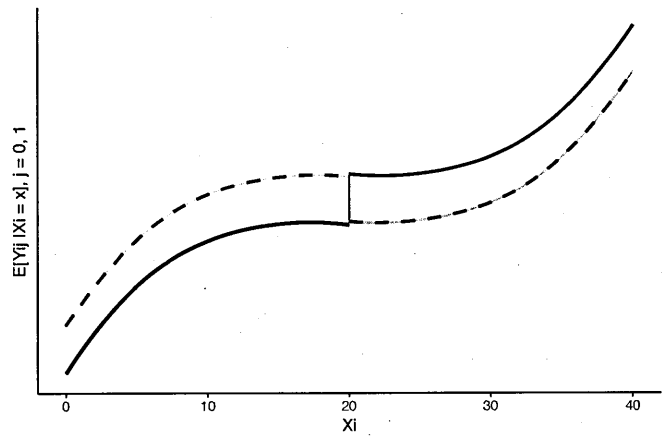

Figure 1.4: Fonction de régression de la variable réponse contrefactuelle (ligne pointillée) et observée (ligne solide)

figure 1.3, une probabilité nulle pour les $X_i < x_0$ et une probabilité de 1 pour les $X_i \geq x_0$. Dans la figure 1.4, les lignes pointillées représentent l'espérance conditionnelle de la variable réponse contrefactuelle : $E[Y_{ij}|X_i = x], \quad j = 0, 1$. Les lignes solides représentent l'espérance conditionnelle de la variable réponse observée : $E[Y_i|X_i = x]$. Ainsi, nous observons que la discontinuité présente dans l'affectation du traitement entraîne une discontinuité dans l'espérance conditionnelle de la variable réponse observée au point x_0 : la mesure de cette discontinuité correspond à l'effet causal moyen du traitement pour les individus pour lesquels $X_i = x_0$.

Nous définissons, formellement, l'effet moyen du traitement au point de discontinuité x_0 comme étant :

$$\begin{aligned} EMT_{x_0} &= E[Y_{i1} - Y_{i0}|X_i = x_0] \\ &= E[Y_{i1}|X_i = x_0] - E[Y_{i0}|X_i = x_0]. \end{aligned} \tag{2}$$

Pour parvenir à identifier et estimer l'effet causal du traitement, le modèle RDS s'appuie sur la comparaison de la variable réponse autour du point de discontinuité :

$$\lim_{x \downarrow x_0} E[Y_i|X_i = x] - \lim_{x \uparrow x_0} E[Y_i|X_i = x]. \tag{3}$$

Les hypothèses d'échangeabilité conditionnelle et de continuité sont nécessaires à l'application du modèle RDS et à l'établissement de l'égalité entre les équations (2) et (3).

H- 1 *Échangeabilité conditionnelle*

L'hypothèse d'échangeabilité conditionnelle se définit comme :

$$Y_{i1}, Y_{i0} \perp\!\!\!\perp T_i|X_i.$$

Dans le cadre du modèle RDS, l'hypothèse d'échangeabilité conditionnelle est nécessairement respectée puisque conditionnellement à la variable d'affectation, le traitement est une fonction déterministe de celle-ci. Or, on ne peut explicitement l'utiliser étant donné que, dans le modèle RDS, l'hypothèse de positivité n'est pas respectée. Cette dernière stipule que chaque sujet a une probabilité strictement positive d'être traité et non traité, ce qui se traduit par l'inégalité suivante :

$$0 < P(T_i = 1|X_i = x) < 1.$$

Or, pour toutes les valeurs de X_i la probabilité de recevoir ou non le traitement, soit le score de propension, est égale à 1 ou 0. En effet, dans le contexte du RDS, un individu avec $X_i < x_0$ ne pourra jamais être exposé au traitement, tandis qu'un individu avec $X_i \geq x_0$ ne pourra jamais être non exposé. Ainsi, l'hypothèse de positivité ne peut être respectée.

Si l'hypothèse d'échangeabilité conditionnelle pouvait être utilisée explicitement dans le modèle RDS, on pourrait alors directement identifier l'EMT au point de discontinuité :

$$\begin{aligned} EMT_{x_0} &= E[Y_{i1}|X_i = x_0] - E[Y_{i0}|X_i = x_0] \\ &= E[Y_{i1}|T_i = 1, X_i = x_0] - E[Y_{i0}|T_i = 0, X_i = x_0] \\ &= E[Y_i|T_i = 1, X_i = x_0] - E[Y_i|T_i = 0, X_i = x_0]. \end{aligned}$$

Or, le non-respect de l'hypothèse de positivité implique que la dernière égalité est non définie puisque nous n'observons aucun individu non traité tel que $X_i = x_0$. Ainsi, pour identifier l'EMT au point de discontinuité, nous avons besoin d'une condition supplémentaire (continuité) décrite comme suit.

H- 2 *Continuité*

L'hypothèse de continuité stipule que :

$$E[Y_{i1}|X = x] \quad \text{et} \quad E[Y_{i0}|X = x] \quad \text{sont continues en } x \text{ et donc en } x_0.$$

La discontinuité de l'espérance conditionnelle de la variable réponse, Y_i, autour du seuil permet l'estimation de l'effet causal. En effet, l'hypothèse de continuité garantit que seule la variable indicatrice du traitement, T_i, est responsable de cette discontinuité (Geneletti *et al.*, 2015). Il existe une version plus générale de l'hypothèse de continuité, qui consiste à émettre les mêmes conditions, mais pour la fonction de répartition des variables contrefactuelles conditionnelle à la variable d'affectation (Imbens et Lemieux, 2008).

Pour les sujets non traités, les hypothèses de continuité et d'échangeabilité conditionnelle impliquent que :

$$\begin{aligned} E[Y_{i0}|X_i = x_0] &= \lim_{x \uparrow x_0} E[Y_{i0}|X_i = x] \quad \textit{(continuité)} \\ &= \lim_{x \uparrow x_0} E[Y_{i0}|T_i = 0, X_i = x] \quad \textit{(échangeabilité conditionnelle)} \\ &= \lim_{x \uparrow x_0} E[Y_i|X_i = x] \quad \textit{(hypothèse de consistance).} \end{aligned}$$

Cette même logique s'applique aussi pour les sujets sous traitement.

Nous pouvons donc identifier l'effet moyen du traitement (EMT) au point de discontinuité comme suit :

$$\begin{aligned} EMT_{x_0} &= E[Y_{i1}|X_i = x_0] - E[Y_{i0}|X_i = x_0] \\ &= \lim_{x \downarrow x_0} E[Y_i|X_i = x] - \lim_{x \uparrow x_0} E[Y_i|X_i = x]. \end{aligned} \tag{3}$$

Pour compléter l'élaboration du modèle RD, il est important de faire la distinction entre le choix du modèle effectué pour la modélisation des données et le choix qui devrait être fait selon le contexte de l'étude. Nous pouvons ajuster un des deux modèles RD principaux aux données à partir des informations a priori sur l'étude,

mais le contexte réel de l'étude peut indiquer un choix contraire. Par exemple, conformément aux indications du NICE, le modèle RDS devrait être ajusté aux données. En réalité, certains professionnels de la santé ne suivent pas strictement les indications ; ils prescrivent des statines aux patients qui ne devraient pas en recevoir. Ainsi, bien que le contexte théorique indique que le choix du RDS serait approprié, le contexte réel de l'étude indique plutôt le modèle RDF.

Pour le reste de ce mémoire, nous supposons que les sujets du modèle RDS sont en parfaite conformité avec la règle de décision ; ils adhèrent strictement et seulement au traitement auquel ils ont été assignés.

1.3 Modèle RD "sharp" et l'estimation de l'EMT

1.3.1 Régression linéaire simple

Dans le but de simplifier la notation, nous définissons les termes suivants :

$$m^{-}(x_0) = \lim_{x \uparrow x_0} E[Y_i | X_i = x]$$

$$m^{+}(x_0) = \lim_{x \downarrow x_0} E[Y_i | X_i = x],$$

ainsi nous pouvons réécrire l'effet moyen du traitement au point de discontinuité comme :

$$EMT_{x_0} = m^{+}(x_0) - m^{-}(x_0).$$

Nous introduisons également un nouveau terme, soit h, la largeur de bande du modèle RD, tel que nous limitons l'estimation de EMT_{x_0} aux valeurs de la variable d'affectation X_i incluses dans l'intervalle $(x_0 - h; x_0 + h)$.

Pour estimer l'effet moyen du traitement au point de discontinuité, il est pertinent de commencer par un procédé non paramétrique : la régression par noyau ou "kernel" (Cerulli, 2015). L'idée générale est d'estimer une moyenne pondérée par

noyau, de part et d'autre du point de discontinuité, pour ensuite différencier ces estimations pour une valeur de largeur de bande fixe. Il s'en suit que l'estimation des deux fonctions de régression $(m^+(x_0), m^-(x_0))$ est :

$$\hat{m}^+(x_0) = \frac{\sum_{i\in D} Y_i \cdot N\left(\frac{X_i - x_0}{h}\right)}{\sum_{i\in D} N\left(\frac{X_i - x_0}{h}\right)} \quad \text{et} \quad \hat{m}^-(x_0) = \frac{\sum_{i\in G} Y_i \cdot N\left(\frac{X_i - x_0}{h}\right)}{\sum_{i\in G} N\left(\frac{X_i - x_0}{h}\right)},$$

où le noyau, $N(\cdot)$, est une fonction positive intégrable à 1, $D = \{i : X_i \geq x_0\}$ et $G = \{i : X_i < x_0\}$. Ainsi nous estimons l'EMT_{x_0} :

$$\widehat{EMT}_{x_0} = \frac{\sum_{i\in D} Y_i \cdot N\left(\frac{X_i - x_0}{h}\right)}{\sum_{i\in D} N\left(\frac{X_i - x_0}{h}\right)} - \frac{\sum_{i\in G} Y_i \cdot N\left(\frac{X_i - x_0}{h}\right)}{\sum_{i\in G} N\left(\frac{X_i - x_0}{h}\right)}. \tag{4}$$

Dans le cas hypothétique où nous utilisons un noyau uniforme sur l'intervalle $(-1, 1)$ tel que :

$$N(z) = \begin{cases} \frac{1}{2} & \text{si } -1 < z < 1 \\ 0 & \text{sinon,} \end{cases}$$

alors l'équation (4) donne :

$$\begin{aligned} \widehat{EMT}_{x_0} &= \frac{\sum_{i\in D} Y_i \cdot \mathbf{1}\{x_0 \leq X_i \leq x_0 + h\}}{\sum_{i\in D} \mathbf{1}\{x_0 \leq X_i \leq x_0 + h\}} - \frac{\sum_{i\in G} Y_i \cdot \mathbf{1}\{x_0 - h \leq X_i < x_0\}}{\sum_{i\in G} \mathbf{1}\{x_0 - h \leq X_i < x_0\}} \qquad (5) \\ &= \overline{Y}_{D,h} - \overline{Y}_{G,h}, \end{aligned}$$

où $\mathbf{1}\{a\}$ est une fonction indicatrice égale à un lorsque a est vraie et zéro sinon. Ainsi, l'utilisation d'une fonction noyau uniforme comme dans l'équation (5) nous

donne un contraste entre la moyenne des variables réponses de part et d'autre du seuil x_0, en tenant compte seulement des observations incluses dans l'intervalle $[x_0 - h; x_0 + h]$.

Dans le cadre de l'estimation de l'effet moyen du traitement, nous nous intéressons à la fonction de régression localisée à un point précis (point de discontinuité), qui est également un point limite. Or l'estimateur (5) présente un problème quant à la gestion de ce type de points. En effet, la régression par noyau fournit des estimateurs dont le comportement est problématique, puisque la vitesse de convergence est moins rapide aux points de discontinuité qu'aux points intérieurs (Imbens et Lemieux, 2008). Ainsi, il peut être démontré (Cerulli, 2015);(Porter, 2003) que le biais asymptotique de l'estimateur par noyau aux points limites est fonction linéaire de la largeur de bande (h), tandis qu'il est fonction de la largeur de bande élevée au carré (h^2) aux points intérieurs.

Nous utilisons donc comme alternative la régression linéaire simple qui tient compte du comportement biaisé des espérances conditionnelles de part et d'autre de x_0 et détient des propriétés fondamentales de réduction du biais dans le cadre du modèle RD (Black *et al.*, 2003). Le processus consiste à ajuster deux fonctions de régression linéaire de part et d'autre du seuil x_0 par la méthode des moindres carrés ordinaires (MCO) :

$$Y_i = \alpha_G + \beta_G(X_i - x_0) + \varepsilon_{G,i} \tag{6}$$

$$Y_i = \alpha_D + \beta_D(X_i - x_0) + \varepsilon_{D,i}, \tag{7}$$

où les erreurs aléatoires, $\varepsilon_{G,i}$ et $\varepsilon_{D,i}$, sont indépendamment distribuées, de variance constante et centrées à zéro. Les fonctions de régressions (6) et(7) se limitent aux observations qui appartiennent uniquement à l'intervalle $[x_0 - h; x_0)$ et $(x_0; x_0 + h]$ respectivement. On implante la méthode des MCO en minimisant la somme des erreurs au carré :

$$\min_{\alpha_G,\beta_G} \sum_{i:x_0-h\leq X_i<x_0} (Y_i - \alpha_G - \beta_G(X_i - x_0))^2$$
$$\min_{\alpha_D,\beta_D} \sum_{i:x_0<X_i\leq x_0+h} (Y_i - \alpha_D - \beta_D(X_i - x_0))^2.$$

Ainsi nous obtenons :

$$\hat{m}^-(x) = \hat{\alpha}_G + \hat{\beta}_G(X_i - x_0)$$
$$\hat{m}^+(x) = \hat{\alpha}_D + \hat{\beta}_D(X_i - x_0).$$

Or dans le cadre du RDS, nous recherchons ces estimations au point de discontinuité x_0, tel que :

$$\hat{m}^-(x_0) = \hat{\alpha}_G$$
$$\hat{m}^+(x_0) = \hat{\alpha}_D.$$

Ceci nous permet donc de trouver l'estimation de l'effet moyen du traitement au point de discontinuité :

$$\begin{aligned} \widehat{EMT}_{x_0} &= \hat{m}^+(x_0) - \hat{m}^-(x_0) \\ &= \hat{\alpha}_D - \hat{\alpha}_G. \end{aligned} \tag{8}$$

Supposons que les erreurs sont indépendamment et identiquement distribuées selon :

$$\varepsilon_{j,i} \sim N(0, \sigma_j^2) \quad j = G, D.$$

Il s'en suit que la variable réponse Y_i est aussi indépendamment distribuée selon une loi normale. Sous ces hypothèses gaussiennes, nous pouvons utiliser la méthode du maximum de vraisemblance pour obtenir les estimateurs des paramètres de régression qui coïncident avec ceux des MCO. Ainsi, nous obtenons les estimateurs suivants pour β_G et β_D :

$$\hat{\beta}_G = \frac{\sum_{i\in G}((x_i - x_0) - \bar{x}_G^c)(Y_i - \overline{Y}_G)}{\sum_{i\in G}((x_i - x_0) - \bar{x}_G^c)^2}$$
$$= \frac{\sum_{i\in G}(x_i - \bar{x}_G)(Y_i - \overline{Y}_G)}{\sum_{i\in G}(x_i - \bar{x})^2},$$

où $\bar{x}_G^c = \frac{1}{n_G}\sum_{i\in G}(x_i - x_0)$, $\bar{x}_G = \frac{1}{n_G}\sum_{i\in G} x_i$ et $\overline{Y}_G = \frac{1}{n_G}\sum_{i\in G} Y_i$.

$$\hat{\beta}_D = \frac{\sum_{i\in D}((x_i - x_0) - \bar{x}_D^c)(Y_i - \overline{Y}_D)}{\sum_{i\in D}((x_i - x_0) - \bar{x}_D^c)^2}$$
$$= \frac{\sum_{i\in D}(x_i - \bar{x}_D)(Y_i - \overline{Y}_D)}{\sum_{i\in D}(x_i - \bar{x})^2},$$

où $\bar{x}_D^c = \frac{1}{n_D}\sum_{i\in D}(x_i - x_0)$, $\bar{x}_D = \frac{1}{n_D}\sum_{i\in D} x_i$ et $\overline{Y}_D = \frac{1}{n_D}\sum_{i\in D} Y_i$.

De même, nous obtenons les estimateurs des paramètres α_G et α_D :

$$\hat{\alpha}_G = \overline{Y}_G - \hat{\beta}_G \bar{x}_G^c \tag{9}$$
$$= \overline{Y}_G - \frac{\bar{x}_G^c}{\sum_{i\in G}(x_i - \bar{x}_G)^2}\left(\sum_{i\in G}(x_i - \bar{x}_G)(Y_i - \overline{Y}_G)\right)$$
$$= \overline{Y}_G - \frac{\bar{x}_G^c}{\sum_{i\in G}(x_i - \bar{x}_G)^2}\left(\sum_{i\in G} x_i Y_i - n_G \bar{x}_G \overline{Y}_G\right).$$

Nous obtenons un résultat analogue pour les $i \in D$:

$$\hat{\alpha}_D = \overline{Y}_D - \frac{\bar{x}_D^c}{\sum_{i\in D}(x_i - \bar{x}_D)^2}\left(\sum_{i\in D} x_i Y_i - n_D \bar{x}_D \overline{Y}_D\right). \tag{10}$$

Il nous est donc possible d'estimer l'effet moyen du traitement au point de discontinuité en utilisant l'équation (8) et les paramètres estimés $\hat{\alpha}_G$ et $\hat{\alpha}_D$.

Il est pertinent de mentionner qu'il n'est pas strictement nécessaire d'effectuer deux régressions séparément pour estimer l'effet moyen du traitement au point de discontinuité. En effet, en combinant les fonctions de régressions (6) et (7) en une seule fonction de régression linéaire et en minimisant les erreurs au carré de celle-ci (Cerulli, 2015), nous pouvons estimer l'effet moyen du traitement identique à (8) :

$$\min_{\alpha_j,\beta_j,\tau} \sum_{i\in I} (Y_i - \alpha_G - \beta_G(X_i - x_0) - \tau T_i - (\beta_D - \beta_G)(X_i - x_0)T_i)^2 \quad j = G, D, \quad (11)$$

où $I = \{i : x_0 - h \leq X_i \leq x_0 + h\}$, T_i est la variable indicatrice du traitement et $\tau = EMT_{x_0}$. Notons que cette approche fait l'hypothèse implicite que $\sigma_G^2 = \sigma_D^2 = \sigma^2$, ce qui est raisonnable de présumer dans un voisinage de x_0.

Une inférence statistique classique s'applique sur τ ou $\alpha_D - \alpha_G$ en autant que la bande h soit sélectionnée adéquatement en fonction de la taille échantillonnale. En particulier, le taux de convergence de h vers zéro doit être suffisament élevé à mesure que la taille d'échantillon tend vers l'infini. Dans le cas contraire, il y aura présence d'un biais asymptotique (Cerulli, 2015).

À noter que la majorité des travaux en RD ont été faits dans un contexte continu, d'où l'utilisation, dans cette section, d'un modèle de régression linéaire pour estimer l'effet moyen du traitement ; l'extension se fait facilement dans un cas de réponse binaire.

1.3.2 Approche bayésienne

Dans le cadre de l'estimation de l'effet moyen du traitement selon l'approche bayésienne, nous devons sélectionner des lois a priori adaptées aux paramètres inconnus $(\alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2)$. Le choix des lois a priori s'appuie habituellement sur des présuppositions résultant d'études déjà publiées ou de recherches approfondies dans la littérature ((O'Keeffe et Baio, 2016) ;(Geneletti *et al.*, 2015)).

Nous conservons notre modélisation gaussienne précédente et utilisons les fonctions de régression (6) et (7) en supposant les erreurs communes. Pour les paramètres $(\alpha_G, \beta_G, \alpha_D, \beta_D)$ nous pouvons choisir des lois a priori normales telles que :

$$\alpha_j \sim N(\mu_{\alpha,j}, \sigma^2/\kappa_{\alpha,j}) \quad \text{et} \quad \beta_j \sim N(\mu_{\beta,j}, \sigma^2/\kappa_{\beta,j}) \quad j = G, D, \tag{i}$$

où $\kappa_{\alpha,j}$ et $\kappa_{\beta,j}$ sont des constantes de proportionnalité tel que si σ^2 est élevée, alors la variance a priori de nos paramètres d'intérêt sera aussi élevée. Ainsi, les informations a priori sur (α_j, β_j) sont fonction de la mesure d'échelle des observations, modulée par les coefficients κ, ces derniers pouvant être vu comme un nombre a priori d'observations (Gelman *et al.*, 2014).

De son côté, la variance doit avoir une loi a priori dont les valeurs s'étendent sur l'ensemble ouvert (0,∞), offrant ainsi plusieurs choix possibles dont une loi gamma inverse ou une loi continue uniforme (O'Keeffe et Baio, 2016). Par exemple, nous pouvons choisir une loi a priori gamma inverse telle que :

$$\sigma^2 \sim IG(\lambda, \omega) \quad \lambda > 0, \omega > 0. \tag{ii}$$

Comme mentionné plus haut, nous choisissons les lois a priori selon les infor-

mations dont nous disposons. Prenons le cas de notre exemple introductif sur le traitement par statines. D'après la littérature, nous savons que le niveau LDL de cholestérol (Y_i) a tendance à diminuer lors de la prise de statines et à augmenter en fonction du taux de risque (Geneletti *et al.*, 2015). Cette information peut être incluse dans le choix des hyperparamètres (i.e. les paramètres des lois a priori), tel que nous avons des valeurs plausibles du niveau LDL de cholestérol pour les taux de risque observés.

Dans l'approche bayésienne, l'inférence sur les paramètres inconnus $(\alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2)$ se base sur la loi a posteriori, qui est proportionelle au produit de la vraisemblance et de la loi a priori :

$$\begin{aligned}\rho(\alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2|Y, X) &\propto \rho(Y|X, \alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2) \\ &\quad \times \rho(\alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2|X) \\ &\propto \rho(Y|X, \alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2) \\ &\quad \times \rho(\alpha_G|\sigma^2)\rho(\beta_G|\sigma^2)\rho(\alpha_D|\sigma^2)\rho(\beta_D|\sigma^2)\rho(\sigma^2),\end{aligned} \tag{12}$$

où $\rho(\alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2|X, Y)$ est la loi a posteriori, $\rho(Y|X, \alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2)$ est la fonction de vraisemblance, $\rho(\alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2|X)$ est la loi conjointe a priori des paramètres du modèle et $\rho(\alpha_G|\sigma^2)$, $\rho(\beta_G|\sigma^2)$, $\rho(\alpha_D|\sigma^2)$, $\rho(\beta_D|\sigma^2)$ et $\rho(\sigma^2)$ sont une factorisation de cette loi, où les coefficients de régression sont a priori présumés indépendants conditionnellement à σ^2.

Conditionnellement à la variance, les lois a posteriori des paramètres d'intérêt (α_G, β_G) et (α_D, β_D) sont indépendamment distribués de part et d'autre du point de discontinuité, ainsi nous pouvons réécrire la loi a posteriori (12) selon la factorisation suivante :

$$\begin{aligned}\rho(\alpha_G,\beta_G,\alpha_D,\beta_D,\sigma^2|X,Y) &= \rho(\alpha_G,\beta_G,\alpha_D,\beta_D|Y,X,\sigma^2)\rho(\sigma^2|Y,X) \qquad (13)\\ &= \rho(\alpha_G,\beta_G|Y,X,\sigma^2)\rho(\alpha_D,\beta_D|Y,X,\sigma^2)\rho(\sigma^2|Y,X)\\ &= \rho(\alpha_G,\beta_G|Y_G,X_G,\sigma^2)\rho(\alpha_D,\beta_D|Y_D,X_D,\sigma^2)\rho(\sigma^2|Y,X).\end{aligned}$$

Il s'en suit, qu'en utilisant les lois a priori des paramètres (i) et la loi a priori de la variance (ii), nous obtenons :

$$\begin{aligned}\rho(\alpha_G,\beta_G,\alpha_D,\beta_D,\sigma^2|X,Y) \propto{}& (\sigma^2)^{-n_G/2}\exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_G}(y_i-(\alpha_G+\beta_G x_i))^2\Big)\times\\ &(\sigma^2)^{-n_D/2}\exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_D}(y_i-(\alpha_D+\beta_D x_i))^2\Big)\times\\ &\Big(\frac{\sigma^2}{\kappa_{\alpha,G}}\Big)^{-1/2}\exp\Big(-\frac{\kappa_{\alpha,G}}{2\sigma^2}(\alpha_G-\mu_{\alpha,G})^2\Big)\times\\ &\Big(\frac{\sigma^2}{\kappa_{\beta,G}}\Big)^{-1/2}\exp\Big(-\frac{\kappa_{\beta,G}}{2\sigma^2}(\beta_G-\mu_{\beta,G})^2\Big)\times\\ &\Big(\frac{\sigma^2}{\kappa_{\alpha,D}}\Big)^{-1/2}\exp\Big(-\frac{\kappa_{\alpha,D}}{2\sigma^2}(\alpha_D-\mu_{\alpha,D})^2\Big)\times\\ &\Big(\frac{\sigma^2}{\kappa_{\beta,D}}\Big)^{-1/2}\exp\Big(-\frac{\kappa_{\beta,D}}{2\sigma^2}(\beta_D-\mu_{\beta,D})^2\Big)\times\\ &(\sigma^2)^{-(\lambda+1)}\exp\Big(-\frac{\omega}{\sigma^2}\Big)\\ \propto{}& (\sigma^2)^{-(\frac{n_G+n_D}{2}+2+\lambda)-1}\times\\ &\exp\Big[-\frac{1}{2\sigma^2}\Big(\sum_{i=1}^{n_G}(y_i-(\alpha_G+\beta_G x_i))^2+\kappa_{\alpha,G}(\alpha_G-\mu_{\alpha,G})^2+\kappa_{\beta,G}(\beta_G-\mu_{\beta,G})^2\Big)\Big]\times\\ &\exp\Big[-\frac{1}{2\sigma^2}\Big(\sum_{i=1}^{n_D}(y_i-(\alpha_D+\beta_D x_i))^2+\kappa_{\alpha,D}(\alpha_G-\mu_{\alpha,D})^2+\kappa_{\beta,D}(\beta_D-\mu_{\beta,D})^2\Big)\Big]\times\\ &\exp\Big(-\frac{\omega}{\sigma^2}\Big).\end{aligned}$$

Dans le but de simplifier l'expression de la loi a posteriori précédente, nous commençons par poser les termes suivants :

$$\nu^* = \frac{1}{2}\Big(\sum_{i=1}^{n_G}(y_i - (\alpha_G + \beta_G x_i))^2 + \kappa_{\alpha,G}(\alpha_G - \mu_{\alpha,G})^2 + \kappa_{\beta,G}(\beta_G - \mu_{\beta,G})^2\Big)$$
$$\nu^{**} = \frac{1}{2}\Big(\sum_{i=1}^{n_D}(y_i - (\alpha_D + \beta_D x_i))^2 + \kappa_{\alpha,D}(\alpha_D - \mu_{\alpha,D})^2 + \kappa_{\beta,D}(\beta_D - \mu_{\beta,D})^2\Big)$$

et obtenons :

$$\begin{aligned}\rho(\alpha_G, \beta_G, \alpha_D, \beta_D, \sigma^2|X, Y) &\propto (\sigma^2)^{-(\frac{n_G+n_D}{2}+2+\lambda)-1}\times \\ &\exp\Big(-\frac{\nu^*}{\sigma^2}\Big) \times \exp\Big(-\frac{\nu^{**}}{\sigma^2}\Big)\times \\ &\exp\Big(-\frac{\omega}{\sigma^2}\Big) \\ &\propto (\sigma^2)^{-(\frac{n_G+n_D}{2}+2+\lambda)-1}\exp\Big[-\frac{1}{\sigma^2}\Big(\omega + \nu^* + \nu^{**}\Big)\Big].\end{aligned}$$

Nous poursuivons dans le même ordre d'idée en utilisant les deux termes suivants :

$$\lambda^* = \Big(\frac{n_G + n_D}{2} + 2 + \lambda\Big)$$
$$\omega^* = (\omega + \nu^* + \nu^{**}).$$

Suite à ces simplifications, nous reconnaissons une loi a posteriori inverse gamma pour la variance conditionnellement à (Y, X) :

$$\rho(\sigma^2|X, Y) \propto (\sigma^2)^{-(\lambda^*+1)}\exp\Big(-\frac{\omega^*}{\sigma^2}\Big),$$

tel que :

$$\sigma^2|Y, X \sim IG(\lambda^*, \omega^*).$$

Par ailleurs, considérons le vecteur $\delta_G = (\alpha_G, \beta_G)$, où les lois a priori respectives des paramètres sont $\alpha_G \sim N(\mu_{\alpha,G}, \sigma^2/\kappa_{\alpha,G})$ et $\beta_G \sim N(\mu_{\beta,G}, \sigma^2/\kappa_{\beta,G})$.

Ainsi, nous pouvons définir la loi a priori de notre vecteur de paramètres δ_G :

$$\delta_G \sim N({\delta_G}^0, \sigma^2 {\boldsymbol{\Sigma}_G}^0),$$

où

$${\delta_G}^0 = (\mu_{\alpha,G}; \mu_{\beta,G})$$

et

$${\boldsymbol{\Sigma}_G}^0 = \begin{pmatrix} 1/\kappa_{\alpha,G} & 0 \\ 0 & 1/\kappa_{\beta,G} \end{pmatrix}$$

est la matrice de variance-covariance à une constante multiplicative près.

Compte tenu des résultats de la factorisation en (13), nous pouvons obtenir une loi a posteriori pour notre vecteur de paramètres δ_G en fonction de la variable réponse et des observations appartenant à l'ensemble $G = \{i : X_i < x_0\}$:

$$\delta_G | Y_G, X_G \sim N({\delta_G}^1, \sigma^2 {\boldsymbol{\Sigma}_G}^1),$$

où $\mathbf{X}_G$ est une matrice de dimension $n_G \times 2$:

$$\mathbf{X}_G = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{n_G} \end{pmatrix}$$

et $\mathbf{Y}_G$ une matrice colonne :

$$[y_1 \dots y_{n_G}]^T,$$

tel que ${\delta_G}^1 = ({\mathbf{X}_G}^T\mathbf{X}_G + \mathbf{\Lambda}_G^0)^{-1}({\mathbf{X}_G}^T\mathbf{Y}_G + \mathbf{\Lambda}_G^0{\delta_G}^0)$ et ${\mathbf{\Sigma}_G}^1 = ({\mathbf{X}_G}^T\mathbf{X}_G + \mathbf{\Lambda}_G^0)$ avec ${\mathbf{\Lambda}_G}^0 = ({\mathbf{\Sigma}_G}^0)^{-1}$.

Nous adoptons la même logique pour obtenir une loi a posteriori similaire pour le vecteur des paramètres δ_D conditionnellement à la variable réponse et aux observations appartenant à l'ensemble $D = \{i : X_i \geq x_0\}$, soit :

$$\delta_D|\mathbf{Y}_D, \mathbf{X}_D \sim N({\delta_D}^1, \sigma^2{\mathbf{\Sigma}_D}^1),$$

où $\mathbf{X}_D$ est une matrice de dimension $n_D \times 2$, $\mathbf{Y}_D = [y_1 \dots y_{n_D}]^T$, tel que ${\delta_D}^1 = ({\mathbf{X}_D}^T\mathbf{X}_D + \mathbf{\Lambda}_D^0)^{-1}({\mathbf{X}_D}^T\mathbf{Y}_D + \mathbf{\Lambda}_D^0{\delta_D}^0)$ et ${\mathbf{\Sigma}_D}^1 = ({\mathbf{X}_D}^T\mathbf{X}_D + \mathbf{\Lambda}_D^0)$ avec ${\mathbf{\Lambda}_D}^0 = (\mathbf{\Sigma}_D^0)^{-1}$.

Il s'en suit que les résultats de la factorisation de la loi a posteriori présentée en (13) propose une méthode simple, par simulation, pour effectuer l'inférence a posteriori sur $\tau = \alpha_G - \alpha_D$. Ainsi, nous avons l'algorithme suivant pour m simulations :

Pour $k = 1, \dots, m$:

1. Échantillonner la variance σ_k^2 selon sa loi a posteriori : $\sigma^2|Y, X \sim IG(\lambda^*, \omega^*)$;
2. Échantillonner $\delta_{G,k}$ selon sa loi a posteriori : $\delta_G|Y_G, X_G, \sigma_k^2 \sim N({\delta_G}^1, \sigma_k^2{\mathbf{\Sigma}_G}^1)$;
3. Échantillonner $\delta_{D,k}$ selon sa loi a posteriori : $\delta_D|Y_D, X_D, \sigma_k^2 \sim N({\delta_D}^1, \sigma_k^2{\mathbf{\Sigma}_D}^1)$;
4. Poser $\tau_k = \alpha_{D,k} - \alpha_{G,k}$.

1.4 Réponse binaire

Il existe plusieurs mesures pour estimer l'effet du traitement dans le cas d'une réponse binaire, dont le risque relatif causal (RRC) et le rapport de cotes causal (RCC). L'hypothèse de continuité sur les espérances conditionnelles H2 nous

permet de considérer ces deux mesures dans notre modélisation dichotomique. En effet, elle est suffisante pour identifier les paramètres de régression structuraux à travers les modèles linéaires généralisés (MLG) qui permettent de relier l'espérance conditionnelle au prédicteur linéaire à l'aide de la fonction lien ("link function") (Bor *et al.*, 2014).

Nous utiliserons, comme mesure de l'effet causal, le RRC qui est l'analogue multiplicatif de l'effet moyen du traitement. Celui-ci se définit comme étant le rapport en moyenne de la réponse contrefactuelle exposée sur la réponse contrefactuelle non exposée au point de discontinuité. En utilisant les hypothèses de modélisation du RDS, nous pouvons écrire le risque relatif causal au point x_0 comme suit :

$$\begin{aligned} RRC_{x_0} &= \frac{E[Y_{i1}|X_i = x_0]}{E[Y_{i0}|X_i = x_0]} \\ &= \frac{\lim_{x \downarrow x_0} E[Y_{i1}|X_i = x]}{\lim_{x \uparrow x_0} E[Y_{i0}|X_i = x]} \quad \textit{(continuité)} \\ &= \frac{\lim_{x \downarrow x_0} E[Y_{i1}|T_i = 1, X_i = x]}{\lim_{x \uparrow x_0} E[Y_{i0}|T_i = 0, X_i = x]} \quad \textit{(échangeabilité conditionnelle)} \\ &= \frac{\lim_{x \downarrow x_0} E[Y_i|X_i = x]}{\lim_{x \uparrow x_0} E[Y_i|X_i = x]} \quad \textit{(hypothèse de consistance).} \end{aligned} \tag{14}$$

Nous poursuivons en posant

$$\begin{aligned} p_1 &= E^+[Y_i|X_i = x] \\ &= P^+[Y_i = 1|X_i = x] \quad \text{et} \\ p_0 &= E^-[Y_i|X_i = x] \\ &= P^-[Y_i = 1|X_i = x], \end{aligned}$$

tel que, dans le cas de la réponse binaire, nous pouvons réécrire l'équation (14) de la façon suivante :

$$RRC_{x_0} = \frac{p_1}{p_0}. \tag{15}$$

Pour être en mesure d'estimer le risque relatif causal, nous pouvons soit ajuster les prédicteurs linéaires des fonctions de régressions (6) et(7) séparément ou utiliser le prédicteur linéaire combiné de l'équation (11).

Selon que nous utilisons un modèle de régression logistique ou un modèle de régression log-binomial, pour estimer le RRC, les coefficients des paramètres de régression impliqués sont interprétés différemment. En effet, les deux modèles ont une fonction lien différente, soit le logit pour la régression logistique et le logarithme pour la régression log-binomiale.

Pour le modèle log-binomial, nous reprenons l'équation (11), ainsi :

$$\log\Big(P(Y_i = 1|X_i = x, T_i = t)\Big) = \alpha_G + \beta_G(X_i - x_0) + \tau T_i + (\beta_D - \beta_G)(X_i - x_0)T_i,$$

où nous incluons les individus dans un voisinage de x_0 tel que $X_i \in [x_0 - h; x_0 + h]$.

Rappelons que $T_i = 1$ lorsque $X_i \geq x_0$ et $T_i = 0$ lorsque $X_i < x_0$; dans le cadre du risque relatif causal, nous cherchons la probabilité de succès dans un voisinage infinitésimal du point de discontinuité x_0. Pour la limite à droite, nous avons :

$$\begin{aligned}
\log(p_1) &= \log\Big(P(Y_i = 1|X_i = x_0, T_i = 1)\Big) \\
&= \alpha_G + \tau \times 1 \\
\Leftrightarrow p_1 &= \exp(\alpha_G + \tau) \\
&= \exp(\alpha_D), \quad \text{si } \tau = \alpha_D - \alpha_G.
\end{aligned}$$

Pour la limite à gauche, nous avons :

$$\begin{aligned}\log(p_0) &= \log\Big(P(Y_i = 1|X_i = x_0, T_i = 0)\Big) \\ &= (\alpha_G + \tau \times 0) \\ &= \alpha_G \\ \Leftrightarrow p_0 &= \exp(\alpha_G).\end{aligned}$$

Ainsi la définition du risque relatif causal établie par le modèle log-binomial est :

$$\begin{aligned}RRC_{x_0}^{bin} &= \frac{p_1}{p_0} \\ &= \exp(\alpha_D - \alpha_G) \\ &= \exp(\tau).\end{aligned} \tag{16}$$

Un des avantages du modèle log-binomial est l'estimation directe du RRC par $\exp(\hat{\tau})$, or ce modèle exhibe souvent des problèmes de type échec de convergence ("failed convergence"). En effet, les MLG sont ajustés en maximisant le logarithme de la fonction de vraisemblance et il y a échec de convergence lorsque le processus de maximisation ne permet pas de retrouver le maximum de vraisemblance (Williamson *et al.*, 2013).

Le modèle logistique est une méthode alternative pour modéliser une réponse binaire qui présente moins de problèmes de convergence que le modèle log-binomial. Il existe d'autres méthodes, outre le modèle logistique, pour contourner cette complication : la régression par le modèle de Cox avec variance robuste proposée par (Lee et Chia, 1993) et évaluée par d'autres chercheurs (Diaz-Quijano, 2012) ;(J D Barros et Hirakata, 2003), ainsi que la régression de Poisson modifiée proposée par (Zou, 2004).

Pour le risque relatif causal, l'approche par le modèle logistique est légèrement

moins directe que celle par le modèle log-binomial. Nous posons, tout d'abord, l'équation de régression suivante :

$$\text{logit}\Big(P(Y_i = 1|X_i = x, T_i = t)\Big) = \alpha_G + \beta_G(X_i - x_0) + \tau T_i + (\beta_D - \beta_G)(X_i - x_0)T_i,$$

où si p est une probabilité, alors $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$.

Similairement, nous recherchons l'expression de la probabilité de succès autour du point de discontinuité x_0. Tout d'abord :

$$\begin{aligned}
\text{logit}(p_1) &= \text{logit}\Big(P(Y_i = 1|X_i = x_0, T_i = 1)\Big) \\
&= \alpha_G + \tau \times 1 \\
\Leftrightarrow \frac{p_1}{1-p_1} &= \exp\Big(\alpha_G + (\alpha_D - \alpha_G)\Big) \\
&= \exp(\alpha_D).
\end{aligned}$$

Suite à quelques manipulations algébriques, nous simplifions l'équation précédente :

$$\begin{aligned}
p_1 &= \frac{\exp(\alpha_D)}{1 + \exp(\alpha_D)} \\
&= \frac{1}{1/\exp(\alpha_D) + 1} \\
&= \frac{1}{1 + \exp(-\alpha_D)}.
\end{aligned}$$

En appliquant la même procédure pour p_0, nous obtenons l'équation du risque relatif causal suivant pour le modèle de régression logistique :

$$RRC_{x_0}^{log} = \frac{1/(1+\exp(-\alpha_D))}{1/(1+\exp(-\alpha_G))} = \frac{1+\exp(-\alpha_G)}{1+\exp(-\alpha_D)}. \quad (17)$$

Nous remarquons que l'utilisation du modèle logistique pour l'estimation du RRC fait intervenir les deux paramètres α_D et α_G distinctement, plutôt que leur différence lorsque le modèle log-binomial est utilisé.

Finalement, nous soulignons que le rapport de cotes causal du modèle logistique peut être vu comme une approximation du risque relatif lorsque la prévalence de la réponse est relativement rare (<10%) (Diaz-Quijano, 2012) ; dans le cas contraire il y aurait une surestimation non négligeable du RRC (J D Barros et Hirakata, 2003). Il s'en suit que, dans le cas du TDAH, nous pouvons interpréter le τ du modèle logistique (RCC) comme un risque relatif causal puisque ce trouble est relativement rare. Mentionnons que les modèles introduits précédemment peuvent être implantés de façon bayésienne ou fréquentiste.

CHAPITRE II

ÉTUDE DE SIMULATION

Nous commençons par définir notre algorithme de simulation principal, puis nous effectuons des analyses de l'impact du changement de taille d'échantillon et de largeur de bande sur l'estimation de l'ETM. Finalement, nous analysons l'aptitude de notre modèle en présence de variables explicative et de confusion.

2.1 Objectifs de simulation

Dans cette section, nous présentons tout d'abord un algorithme standard pour simuler notre base de données, répliquant ainsi une RDS adaptée à notre étude de cas. Par la suite, nous faisons une analyse de sensibilité pour vérifier si nos résultats sont affectés par la variation de certains facteurs, tels que la largeur de bande et la taille d'échantillon. Nous examinons aussi l'impact de l'ajout d'une variable explicative sur la précision de l'estimation de l'effet moyen du traitement, ainsi que l'effet produit par la présence de confusion dans une RDS.

2.2 Simulation standard d'un jeu de données pour une RDS

Plusieurs études ont conclu qu'il existait un lien entre l'âge d'entrée à l'école et l'incidence d'être diagnostiqué d'un trouble déficitaire de l'attention avec ou sans

hyperactivité (TDAH)((L Morrow *et al.*, 2012), (E Elder, 2010), (N Evans *et al.*, 2010), (Halldner *et al.*, 2014) et (Krabbe *et al.*, 2014)). Entres autres, (N Evans *et al.*, 2010) ont rapporté que les enfants nés juste après la date éligible d'entrée ont une incidence de diagnostic et de traitement TDAH significativement plus basse que les enfants nés juste avant celle-ci. (L Morrow *et al.*, 2012) ont obtenu des résultats similaires qui démontrent un effet de l'âge relatif sur le diagnostic et le traitement de TDAH, soulevant une inquiétude quant aux conséquences reliées au surdiagnostic et prescriptions excessives. Finalement, (E Elder, 2010) ont déduit que les enfants nés au cours du mois précédent la date limite d'entrée, étant généralement les plus jeunes et les plus immatures dans la classe, reçoivent un plus grand nombre de diagnostics de TDAH que les enfants nés le mois suivant. Ils soulignent également que ceci influence fortement les évaluations des enseignants quant à savoir si l'enfant présente des symptômes de TDAH, suggérant que de nombreux diagnostics peuvent être motivés par la perception des enseignants d'un mauvais comportement chez les plus jeunes.

Ainsi, dans le cas de notre étude de simulation, nous nous questionnons sur la présence d'un effet de l'âge relatif à l'entrée d'un enfant à l'école sur la probabilité d'être diagnostiqué (TDAH). Il s'en suit que nous avons une variable réponse Y_i qui est égale à 1 si l'enfant reçoit le diagnostic de TDAH et de 0 dans le cas contraire, tel que :

$$logit\Big(P(Y_i = 1|X_i = x, T_i = t)\Big) = \alpha_G + \beta_G(X_i - x_0) + \tau T_i + (\beta_D - \beta_G)(X_i - x_0)T_i, \quad (18)$$

où τ mesure l'impact d'être né après la date éligible d'entrée sur la probabilité de recevoir un diagnostic de TDAH.

Selon la Commission scolaire de Montréal[1], l'enfant doit être âgé de 5 ans, en date du 30 septembre, pour être admis à l'école préscolaire ; cette date éligible d'entrée représente donc notre point de discontinuité. Il s'en suit que nous incluons dans notre simulation les enfants dont la date de naissance se situe à 183 jours (selon une année bissextile : 366/2) de part et d'autre de la date éligible d'entrée, ce qui constitue notre largeur de bande que nous ferons varier lors de notre étude de sensibilité.

Ainsi, nous avons l'algorithme suivant pour simuler m jeux de données :

1. Simuler aléatoirement une date de naissance (jour, mois) $\mathbf{X} = (X_1, \ldots, X_n)^T$ pour n enfants. Pour avoir un point de discontinuité centré à l'origine, nous introduisons $X_i^c = X_i - x_0$ et nous conservons uniquement les enfants dont la date de naissance, X_i^c, se situe à l'intérieur de l'intervalle de bande $[-183 - h; 183 + h]$.
2. Définir la variable indicatrice du traitement T_i. Celle-ci est égale à 1 si l'enfant est né après la date éligible d'entrée et 0 sinon, ce qui équivaut à :
$$T_i = \begin{cases} 1 & \text{si } X_i^c \geq 0 \\ 0 & \text{si } X_i^c < 0. \end{cases}$$
3. Tirer aléatoirement la réponse $Y_i (i = 1, \ldots, n)$ d'une loi de Bernoulli avec une probabilité de succès donnée par le modèle logistique (18).

   3.1. Soit g, une combinaison linéaire de nos prédicteurs, telle que :
   $$g = \alpha_G + \beta_G X_i^c + \tau T_i + (\beta_D - \beta_G) X_i^c T_i,$$
   où $\tau = \alpha_D - \alpha_G$ et les coefficients de régression $(\alpha_D, \alpha_G, \beta_D, \beta_G)$ sont fixés selon des informations a priori tirées de la littérature. Ainsi :

1. http://csdm.ca/parents-eleves/admission/

$$g_1 = \alpha_D + \beta_D X_i^c, \quad \text{si } T_i = 1$$
$$g_0 = \alpha_G + \beta_G X_i^c, \quad \text{si } T_i = 0.$$

3.2. Nous pouvons, désormais, calculer les probabilités de succès, selon le modèle logistique suivant :

$$p_1 = \frac{\exp(g_1)}{(1+\exp(g_1))}, \quad \text{si } T_i = 1$$
$$p_0 = \frac{\exp(g_0)}{(1+\exp(g_0))}, \quad \text{si } T_i = 0.$$

Ainsi, nous générons aléatoirement la variable réponse binaire selon les lois de Bernoulli respectives :

$$Y_i \sim \begin{cases} Ber(pr_1) & \text{si } T_i = 1 \\ Ber(pr_0) & \text{si } T_i = 0. \end{cases}$$

4. Répéter les étapes précédentes m fois, créant ainsi m bases de données simulées.

Il s'en suit que nous concevons une fonction permettant de simuler nos m jeux de données et, pour chacun d'entres eux, ajustons un modèle de régression logistique des Y_i simulées en fonction de la variable d'affectation, la variable indicatrice du traitement et leur interaction (18). Nous estimons ainsi m fois le paramètre d'intérêt, τ, qui représente l'effet du traitement.

Pour réussir à exécuter notre fonction, nous devons fixer les coefficients de régression suivants : $\alpha_D, \beta_D, \alpha_G, \beta_G$. Rappelons qu'en assignant une valeur prédéfinie à τ, nous appliquons une contrainte sur α_D et α_G, qui se traduit selon l'équation suivante : $\tau = \alpha_D - \alpha_G$.

On détermine la taille d'effet d'après les informations provenant d'études sur l'effet de l'âge relatif d'entrée de l'enfant à l'école sur le diagnostic de TDAH. Ainsi, plus

la date de naissance de l'enfant s'éloigne de la date éligible d'entrée, plus nous avons une diminution de la probabilité de recevoir un diagnostic.

Pour estimer l'effet moyen du traitement, (N Evans *et al.*, 2010) et (E Elder, 2010) utilisent, dans leurs études sur le lien entre l'âge relatif de l'enfant et l'incidence de diagnostic de TDAH, un modèle linéaire avec une variable réponse binaire. Par conséquent, leur coefficient de régression associé au traitement s'interprète comme étant une différence de risques (DR). Celle-ci se définit comme étant la différence entre la probabilité d'être diagnostiqué TDAH selon que l'enfant a une entrée tardive ou hâtive, donc selon la position de la date de naissance de l'enfant par rapport au point de discontinuité. Les coefficients variant entre -0.02 à -0.05, nous choisissons arbitrairement une différence de risques de -0.03 sur laquelle nous allons nous inspirer pour spécifier les valeurs de nos paramètres.

Il s'en suit que si nous posons :

$$DR = P(Y_i = 1|X_i = x_0, T_i = 1) - P(Y_i = 1|X_i = x_0, T_i = 0)$$

$$-0.03 = 0.04 - 0.07,$$

alors le rapport de cotes est égal à :

$$\begin{aligned} RCC &= \frac{P(Y_i = 1|X_i = x_0, T_i = 1)/(1 - P(Y_i = 1|X_i = x_0, T_i = 1))}{P(Y_i = 1|X_i = x_0, T_i = 0)/(1 - P(Y_i = 1|X_i = x_0, T_i = 0))} \\ &= \frac{0.04/(1 - 0.04)}{0.07/(1 - 0.07)} \\ &= \frac{0.04/0.96}{0.07/0.93} \\ &= 0.554. \end{aligned}$$

Rappelons que lorsqu'on utilise un modèle de régression logistique, les probabilités conditionnelles autour du point de discontinuité $P(Y_i = 1|X_i = x_0, T_i = t)$, se calculent comme suit :

$$p_1 = \frac{1}{1+\exp(-\alpha_D)} \quad \text{et} \quad p_0 = \frac{1}{1+\exp(-\alpha_G)}.$$

Ainsi, nous devons résoudre le système d'équations suivant, afin de retrouver les paramètres qui vont correspondre à un effet du traitement estimé de l'ordre de $\tau = log(0.554) = -0.591$:

$$\begin{aligned} 0.04 &= \frac{1}{1+\exp(-\alpha_D)} \\ \Leftrightarrow \alpha_D &= -log(0.96/0.04) \\ &= -3.178 \end{aligned}$$

et

$$\begin{aligned} 0.07 &= \frac{1}{1+\exp(-\alpha_G)} \\ \Leftrightarrow \alpha_G &= -log(0.93/0.07) \\ &= -2.587, \end{aligned}$$

vérifiant que :

$$\begin{aligned} \tau &= \alpha_D - \alpha_G \\ &= -3.178 - (-2.5867) \\ &= -0.591. \end{aligned}$$

On s'attend à la présence d'une circularité dans nos données ; d'une année à l'autre, la proportion d'enfants recevant un diagnostic de TDHA va croître plus leur date de naissance s'approche de la date éligible d'entrée à l'école, pour ensuite diminuer plus leur date de naissance s'éloigne du point de discontinuité. Ainsi, nous fixons des pentes identiques et positives, tel que le logarithme de la cote de recevoir

un diagnostic de TDAH s'accroît faiblement plus la date de naissance de l'enfant s'éloigne positivement du point de discontinuité : $\beta_D = 0.00162$ et $\beta_G = 0.00162$ (voir calculs détaillés dans l'Annexe A).

La figure 2.1 suivante est un exemple d'une simulation d'un jeux de données contenant 20 000 individus où l'axe des ordonnées est le logit de la réponse simulée. Comme notre modèle de simulation est généré selon une régression logistique, la réponse simulée est exprimée sur le *logit* de la probabilité que l'enfant soit diagnostiqué TDAH ($Y_i = 1$). Nous avons partitionné l'année, allant de -183 à 183, en utilisant 60 intervalles de temps. Ainsi, au lieu de calculer la probabilité empirique de la réponse diagnostiquée à chaque temps, X_i, nous l'avons calculé pour chaque intervalle de temps ; chaque unité sur l'axe des abscisses représente un intervalle de temps. Par exemple,

$$\widehat{\Pr}(Y_i = 1 | X_i \in [-128, -122)) = \left(\frac{\text{nbre de } Y_i = 1 | X_i \in [-128, -122)}{\text{nbre de } X_i \in [-128, -122)} \right),$$

il s'en suit que chaque point dans la figure représente ces probabilités estimées.

La ligne solide représente la droite de régression selon que les individus sont situés à gauche du point de discontinuité, tandis que la ligne pointillée fait référence à la droite de régression pour ceux situés à droite[2].

2.2.1 Impact de la taille échantillonnale

Nous débutons notre étude de simulation en regardant l'effet des trois tailles d'échantillon n suivantes sur le biais, la variance et l'erreur quadratique moyenne (EQM) de l'estimateur de l'effet du traitement (τ) : 250, 500 et 1000. Ainsi, pour

---

2. Selon la valeur des paramètres $(\alpha_D, \alpha_G, \beta_D, \beta_G)$, nous avons $E[Y_{i,D}|X_i] = -3.178 + 0.00162 \cdot X_i$ et $E[Y_{i,G}|X_i] = -2.5867 + 0.00162 \cdot X_i$

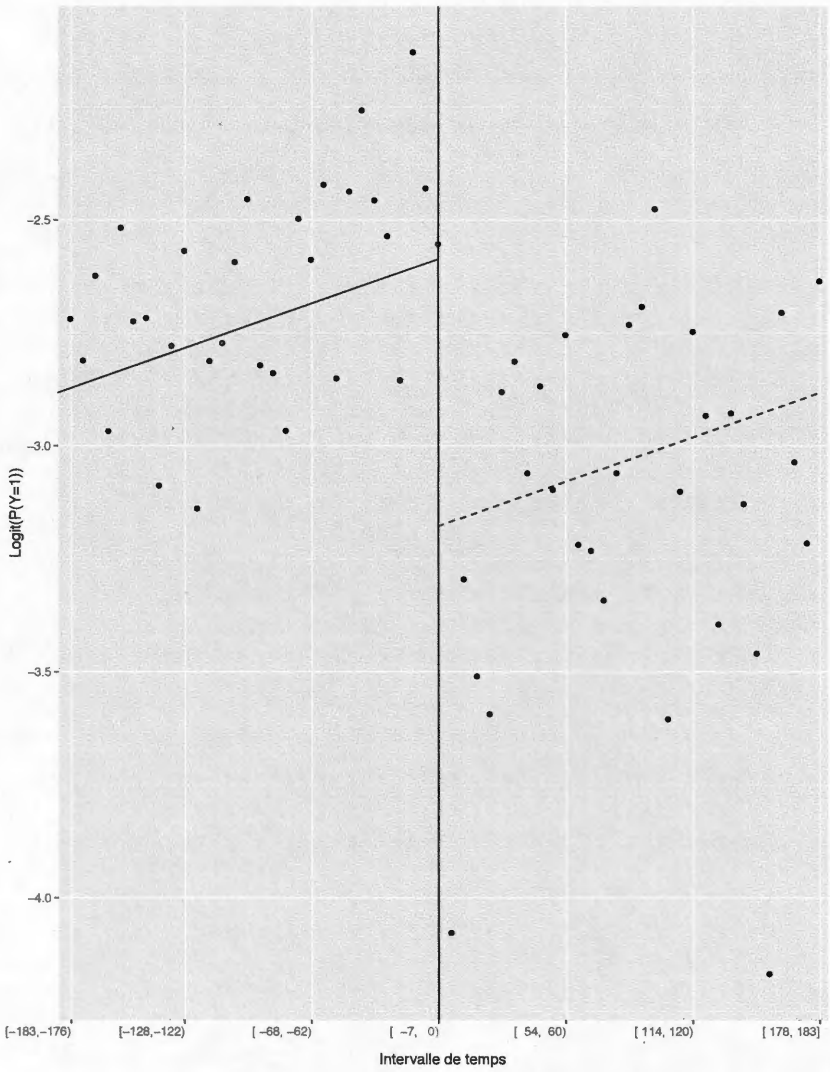

Figure 2.1: Exemple d'un jeux de données pour 20 000 individus avec droite de régression selon que l'individu soit né avant (ligne solide) ou après (ligne pointillée) le point de discontinuité

chaque taille d'échantillon, ces trois mesures sont calculées selon une taille de simulation et une largeur de bande fixes, soient 1000 et 183 respectivement.

Les résultats de cette étude sont résumés dans le tableau 2.1, où le RCEQM représente la racine carrée de l'erreur quadratique moyenne et l'ETM est l'erreur type moyenne de l'estimateur. Le calcul de l'écart-type s'est fait selon l'équation suivante :

$$\sigma = \sqrt{E[(\hat{\tau} - E(\hat{\tau}))^2]}.$$

En premier lieu, nous pouvons remarquer que notre estimateur a un biais estimé pratiquement nul, ce qui n'est pas surprenant puisqu'on ajuste le même modèle utilisé pour simuler nos données. Pour ce qui est de l'écart-type, nous remarquons que celui-ci diminue au fur et à mesure que la taille d'échantillon augmente. Nous constatons aussi que les valeurs estimées de l'ETM coincident avec celles de l'écart-type ; la méthode d'estimation de la variabilité de $\hat{\tau}$ est donc juste.

En général, puisque l'erreur quadratique moyenne est fonction de la variance et du biais, ses fluctuations vont dépendre de la direction que prend la tendance de ces deux mesures. L'équation de l'EQM suivante traduit la relation qui existe entre la variance et le biais :

$$EQM(\hat{\tau}) = Var(\hat{\tau}) + Biais(\hat{\tau}, \tau)^2. \qquad \text{(iii)}$$

Or, comme le biais de l'estimateur est faible et que l'écart-type diminue quand la taille d'échantillon augmente, la RCEQM décroît au même ordre que l'écart-type. Nous concluons donc que la précision de notre estimateur sans biais augmente avec la taille d'échantillon.

Tableau 2.1: Impact de la taille échantillonnale pour une simulation Monte-Carlo avec 1000 réplications, $\tau = -0.5913$

Taille d'échantillon	Moyenne	Biais	Écart-type	RCEQM	ETM
250	−0.603	−0.011	0.580	0.581	0.575
500	−0.595	−0.003	0.396	0.396	0.401
1000	−0.584	0.008	0.280	0.280	0.282

2.2.2 Impact de la largeur de bande

Nous poursuivons notre étude en utilisant les trois h suivantes : 30, 92, et 183 jours en gardant le nombre de réplications de jeux de données et leurs tailles fixes (1000 et 500 respectivement). Les résultats de cette étude se retrouvent dans la table 2.2.

Puisque notre modèle de simulation n'inclut pas de variables de confusion, alors la diminution de la fenêtre de temps ne devrait pas influencer le biais ; celui-ci demeure en effet presque nul. De plus, nous constatons que l'écart-type de notre estimateur diminue avec l'augmentation de la fenêtre de temps. Ceci n'est pas surprenant puisqu'une petite largeur de bande implique que seulement une partie de nos observations est utilisée pour estimer l'effet du traitement. Comme dans le cas de l'étude sur la taille échantillonnale, puisque le biais est relativement faible et que l'écart-type diminue avec la largeur de bande, la RCEQM décroît au même ordre que ce dernier ; la précision de notre estimateur augmente avec la largeur de la fenêtre de temps. Pour un $h = 30$, nous remarquons que, l'ETM est légèrement plus petite que l'écart-type, ce qui nous indique une faible sous estimation de la variabilité. Pour ce qui est des autres largeurs de bande, les valeurs de l'ETM sont relativement proches de celles des écarts-types.

À noter, puisque notre modèle de simulation ne comprend pas encore de variable de confusion, l'effet de la taille d'échantillon est équivalente à celle des fenêtres de temps (voir Annexe B).

Tableau 2.2: Impact de la largeur de bande pour une simulation Monte-Carlo avec 1000 réplications et une taille échantillonnale de 500, $\tau = -0.5913$

Largeurs de bande	**Moyenne**	**Biais**	**Écart-type**	**RCEQM**	**ETM**
30	−0.594	−0.002	1.178	1.178	1.077
92	−0.581	−0.010	0.569	0.569	0.575
183	−0.595	−0.003	0.396	0.396	0.401

2.3 Impact de l'ajout d'une variable explicative

En pratique, il est assez courant d'ajouter des covariables supplémentaires dans la spécification d'un modèle de régression. Ces variables explicatives peuvent, entre autres, être utilisées pour diminuer le biais échantillonnal et améliorer la précision des estimateurs. Dans le contexte RD, leur présence a très peu d'effet sur l'identification de l'effet moyen du traitement dans un petit voisinage du point de discontinuité. Dans le cas où la fenêtre de temps autour du point de discontinuité est plus importante, un effet pourrait être perçu dans le cas où il y a de la confusion (Imbens et Lemieux, 2008).

Faisons donc l'ajout de la variable du sexe dans notre modèle de simulation, telle que notre équation de régression s'écrit désormais comme suit :

$$\begin{aligned}\text{logit}\Big(P(Y_i = 1|X_i = x, T_i = t, S_i = s)\Big) = \alpha_G + \beta_G(X_i - x_0) + \tau T_i + \\ (\beta_D - \beta_G)(X_i - x_0)T_i + \beta_S S_i,\end{aligned} \tag{19}$$

où S_i est la variable indicatrice du sexe tel que :

$$S_i \sim \begin{cases} 1 & \text{si c'est un garçon} \\ 0 & \text{si c'est une fille.} \end{cases}$$

Il s'en suit que nous préservons notre processus de simulation précédemment défini, mais que nous rajoutons certaines spécifications propres à la variable du sexe. Notre nouveau modèle de simulation permet des ordonnées à l'origine légèrement plus élevées pour les garçons que pour les filles. Ainsi, nous modifions les premières étapes de notre modèle de simulation, tel que :

$$g_1 = \alpha_D + \beta_D X_i^c + \beta_S S_i, \quad \text{si } T_i = 1$$
$$g_0 = \alpha_G + \beta_G X_i^c + \beta_S S_i, \quad \text{si } T_i = 0,$$

où $\beta_S = 0.6$.

Les garçons ont désormais comme ordonnées à l'origine $(\alpha_D + \beta_S)$ et $(\alpha_G + \beta_S)$, tandis que les filles ont (α_D, α_G), toutefois l'effet du traitement (sur le logarithme du rapport de cote) demeure le même dans les deux strates : $\tau = \alpha_D - \alpha_G$.

La table 2.3 nous présente les probabilités calculées au point de discontinuité pour chaque sexe. Pour ce faire, nous avons résolu l'équation (19) avec le β_S sélectionné, selon que l'enfant soit né après le point de discontinuité $(T_i = 1)$ ou avant $(T_i = 0)$. Nous obtenons ainsi les équations de probabilités suivantes :

$$p_1 = \frac{\exp(\alpha_G + \tau + \beta_S S_i)}{(1 + \exp(\alpha_G + \tau + \beta_S S_i))}, \quad \text{si } T_i = 1$$
$$p_0 = \frac{\exp(\alpha_G + \beta_S S_i)}{(1 + \exp(\alpha_G + \beta_S S_i))}, \quad \text{si } T_i = 0.$$

Comme nous nous attendons à ce que les garçons reçoivent un taux de diagnostic TDAH plus élevé que celui des filles ((N Evans *et al.*, 2010),(L Morrow *et al.*, 2012)), le choix du $\beta_S = 0.6$ semble justifié puisque nous retrouvons des valeurs relativement réalistes supportant cette hypothèse.

Tableau 2.3: Probabilités théoriques, au point de discontinuité, de la régression logistique avec la variable de sexe $\beta_S = 0.6$

T_i	**Fille**	**Garçon**
1	0.040	0.071
0	0.070	0.121

Nous poursuivons en simulant des jeux de données avec le terme de sexe et en comparant la précison de notre estimateur $\hat{\tau}$ obtenu d'un modèle qui ajuste ou qui n'ajuste pas pour cette variable. Pour avoir la possibilité de démontrer l'effet de l'ajout d'une variable explicative, nous simulons pour deux forces de β_S différentes. En effet, plus le β_S est grand en valeur absolue, plus on pourrait penser qu'il y a avantage à ajuster pour la variable dans notre modèle de régression. Les résultats présentés dans la table 2.4 sont pour le $\beta_S = 0.6$ et ceux dans la table 2.5 pour un $\beta_S = 1.5$. Notons que nous avons ajouté une nouvelle mesure d'efficacité et de précision de notre modèle de simulation, soit la probabilité de couverture qui nous indique la probabilité que notre procédure d'estimation produit un intervalle contenant la vraie valeur de τ (Shedden, 2019).

Il s'en suit que nous présentons uniquement les résultats pour la taille échantillonnale, puisque, comme mentionné plus haut, l'effet de la taille d'échantillon est équivalent à celle de la fenêtre de temps sans confusion. Nous pouvons tout d'abord remarquer que, dans le cas du β_S élevé, la différence entre l'écart-type et l'ETM est non négligeable, entre autres, pour la plus petite taille d'échantillon.

Le biais, avec ou sans ajustement du terme sexe, est faible pour les deux β_S. Par contre, on peut constater que les moyennes ajustées et celles non ajustées du $\hat{\tau}$ concordent moins pour le β_S élevé que pour le β_S plus petit. Ce comportement s'explique par le phénomène de non-collapsibilité des rapports de cotes du modèle de régression logistique ; l'effet du traitement estimé change lorsque l'on ajuste pour des variables explicatives (Ciolino *et al.*, 2013), (Guo et Geng, 1995).

De plus, on pourrait s'attendre à ce que les écart-types, après ajustement, soient moins élevés, or les résultats présentent une situation contraire ; ils sont plus élevés après ajustement quelle que soit la taille d'échantillon. Il s'avère que, pour les modèles classiques de régression, l'ajustement d'une variable explicative non-confondante augmente la précision de l'estimateur, tandis que pour les modèles logistiques l'ajustement résulte en une perte de précision ((Xing et Xing, 2010),(Ciolino *et al.*, 2013),(Robinson et Jewell, 1991)).

Dans le cas du modèle logistique, certains auteurs ((Xing et Xing, 2010),(Ciolino *et al.*, 2013)) suggèrent toutefois de garder l'ajustement de la variable explicative, car la perte de précision vient avec un gain de puissance. Pour que ce gain soit observé, il faut qu'il existe une différence entre les effets marginaux (non-ajustés) et conditionnels (ajustés) et que ces derniers soient plus éloignés de la valeur de référence. Nous pouvons, en effet, observer ce phénomène dans la table 2.5. Finalement, nous constatons que les probabilités de couverture estimées se situent toutes autour de 94% et 95%, soit relativement proche de la probabilité nominale de confiance 95% comme souhaité.

Tableau 2.4: Impact de l'ajustement sur la variable sexe, en fonction de la taille échantillonnale, pour un effet modéré de la variable ($\beta_S = 0.6$), $\tau = -0.5913$

Taille d'échantillon	**Moyenne**	**Biais**	**Écart-type**	**RCEQM**	**ETM**	**Probabilité de couverture**
Scénario 1 : Sans ajustement de la variable sexe						
250	−0.594	−0.002	0.528	0.528	0.532	0.945
500	−0.598	−0.007	0.383	0.384	0.372	0.946
1000	−0.590	0.001	0.265	0.265	0.262	0.943
Scénario 2 : Avec ajustement de la variable sexe						
250	−0.606	−0.015	0.537	0.537	0.538	0.946
500	−0.607	−0.015	0.388	0.388	0.376	0.944
1000	−0.599	−0.008	0.266	0.266	0.264	0.948

Tableau 2.5: Impact de l'ajustement sur la variable sexe, en fonction de la taille échantillonnale, pour un effet fort de la variable ($\beta_S = 1.5$), $\tau = -0.5913$

Taille d'échantillon	**Moyenne**	**Biais**	**Écart-type**	**RCEQM**	**ETM**	**Probabilité de couverture**
Scénario 1 : Sans ajustement de la variable sexe						
250	−0.601	−0.009	0.749	0.749	0.719	0.942
500	−0.601	−0.010	0.506	0.506	0.498	0.951
1000	−0.586	0.006	0.364	0.364	0.348	0.940
Scénario 2 : Avec ajustement de la variable sexe						
250	−0.625	−0.034	0.759	0.760	0.732	0.942
500	−0.620	−0.029	0.518	0.519	0.506	0.940
1000	−0.602	−0.011	0.369	0.370	0.354	0.938

Les tableaux 2.6 rapportent l'analyse de puissance effectuée pour chaque valeur de β_S utilisée pour générer les données simulées. Pour ce faire, nous avons calculé en moyenne le nombre de fois que l'on rejetait l'hypothèse nulle $\tau = 0$ pour

un niveau fixe de 0.05. Nous obtenons, effectivement, un gain de puissance plus prononcé pour la plus grande valeur de β_S.

Tableau 2.6: Impact de l'ajustement pour la variable de sexe sur la puissance pour une simulation Monte-Carlo avec 1000 réplications, $\tau = -0.5913$

Taille d'échantillon	**Puissance statistique**
Scénario 1 : Sans ajustement de la variable sexe, $\beta_S = 0.6$	
250	0.187
500	0.353
1000	0.616
Scénario 2 : Avec ajustement de la variable sexe, $\beta_S = 0.6$	
250	0.189
500	0.35
1000	0.626
Scénario 1 : Sans ajustement de la variable sexe, $\beta_S = 1.5$	
250	0.17
500	0.315
1000	0.571
Scénario 2 : Avec ajustement de la variable sexe, $\beta_S = 1.5$	
250	0.192
500	0.326
1000	0.586

2.4 Impact de l'ajout d'une variable de confusion

L'objectif de l'analyse de l'impact de la confusion est de démonter l'efficacité du modèle RD quant à fournir un estimateur non biaisé dans un voisinage du point de discontinuité x_0 en présence de variables confondantes.

Pour introduire l'effet d'une variable de confusion dans notre modèle de simulation, nous devons trouver une variable qui a une association commune avec la probabilité d'être diagnostiqué TDAH (la réponse) et la variable d'affectation (voir section 1.1.4). Or, dans le cas de notre étude, imaginer ce qui influencerait à la fois la réponse et la date de naissance s'avère difficile. Par conséquent, nous avons créé un context fictif qui nous donne quand même l'opportunité d'analyser l'effet de la largeur de bande dans un tel contexte RD.

Il s'en suit que nous avons décidé d'utiliser un facteur socio-économique, soit le revenu annuel gagné, comme variable confondante. Celui-ci se divise en deux catégories : moins que 50 000$ et plus de 50 000$. Comme les ressources au système de santé au Québec sont économiquement facilement accessibles, nous déterminons qu'il existe une association négative entre notre facteur socio-économique et la probabilité de diagnostic : ceux dont le statut est relativement confortable ont une probabilité de diagnostic plus faible que les individus dont la situation est plus précaire. Pour ce qui est de l'association avec les dates de naissance des enfants, c'est ici qu'entre en jeu la création d'un contexte de simulation fictif. En effet, nous décidons que les parents dont le revenu est plus faible, auront plus de naissances à une certaine période de l'année (après la date d'entrée éligible), comparativement aux enfants dont la situation économique est aisée. Ainsi, nous avons les associations comme présentées dans la figure 2.2.

Pour traduire notre modèle de simulation, nous avons effectué les étapes suivantes :

1. Simuler aléatoirement la variable du statut économique, $Statut_i$, à deux catégories :

$$0 = \text{moins de } 50\ 000\$$$

$$1 = 50\ 000\$ \text{ et plus},$$

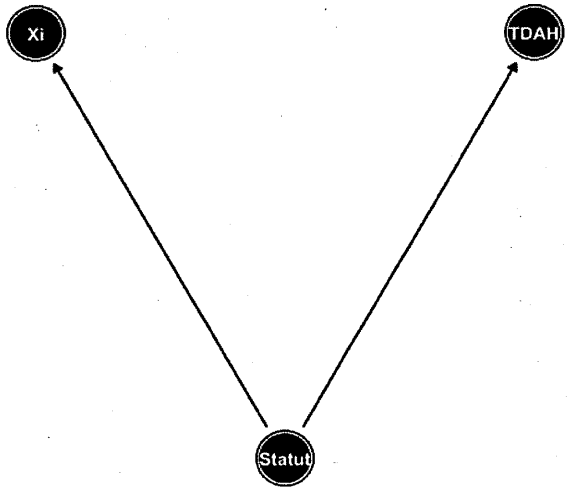

Figure 2.2: Graphique orienté acyclique causal ("DAG") représentant les associations avec la variable de confusion

avec les probabilités respectives de 0.714 et 0.286 [3].

2. Définir la variable d'affectation qui désormais doit inclure la composante d'association avec la variable confondante :

$$X_i^c = U(-183, 183) + \beta_{statutX} Statut_i,$$

où nous tronquons les enfants dont la date de naissance, X_i^c, se situe à l'extérieur de l'intervalle de bande $[-h; h]$.

La figure dans l'Annexe C nous donne un aperçu de la distribution empirique des dates de naissance des enfants en fonction de leur statut économique simulé.

3. Définir la variable indicatrice T_i. Celle-ci est égale à 1 si l'enfant est né après

3. Ces probabilités ont été calculées selon Revenu Québec : `https://www.revenuquebec.ca/fr/salle-de-presse/statistiques/le-revenu-total-des-particuliers/`

la date éligible d'entrée et 0 sinon, ce qui équivaut à :

$$T_i = \begin{cases} 1 & \text{si } X_i^c \geq 0 \\ 0 & \text{si } X_i^c < 0. \end{cases}$$

4. Tirer aléatoirement la réponse $Y_i (i = 1, \ldots, n)$ d'une loi de Bernoulli avec une probabilité de succès donnée par le modèle logistique.

   4.1. Soit g, une combinaison linéaire de nos prédicteurs, telle que :

   $$g = \alpha_G + \beta_G X_i^c + \tau T_i + (\beta_D - \beta_G) X_i^c T_i + \beta_{statutY} Statut_i,$$

   Ainsi :

   $$g_1 = \alpha_D + \beta_D X_i^c + \beta_{statutY} Statut_i, \quad \text{si } T_i = 1$$
   $$g_0 = \alpha_G + \beta_G X_i^c + \beta_{statutY} Statut_i, \quad \text{si } T_i = 0.$$

   4.2. Nous pouvons, désormais, calculer les probabilités de succès, selon le modèle logistique suivant :

   $$pr_1 = \frac{\exp(g_1)}{(1 + \exp(g_1))}, \quad \text{si } T_i = 1$$
   $$pr_0 = \frac{\exp(g_0)}{(1 + \exp(g_0))}, \quad \text{si } T_i = 0.$$

   Ainsi, nous générons aléatoirement la variable de réponse binaire selon les lois de Bernoulli respectives :

   $$Y_i \sim \begin{cases} Ber(pr_1) & \text{si } T_i = 1 \\ Ber(pr_0) & \text{si } T_i = 0. \end{cases}$$

   Nous retrouvons, dans l'Annexe D, un aperçu de la distribution empirique de la probabilité d'être diagnostiqué TDAH en fonction du statut socioéconomique.

5. Répéter les étapes précédentes m fois, créant ainsi m bases de données simulées.

Comme nous tentons de modéliser une situation fictive où il y a de la confusion, nous avons essayé de nombreux choix de β_{statut} qui ne sont pas nécessairement

réalistes quand à l'amplitude de l'effet sur la réponse ou la variable d'affectation. Nous avons ainsi retenus, comme choix finaux, un $\beta_{statutX} = 30$ et un $\beta_{statutY} = -0.2$. De plus, pour faire ressortir un effet caractéristique de confusion nous avons augmenté la taille d'échantillon à 10 000.

Les résultats de l'impact de l'ajout de notre variable de confusion se retrouvent dans la table 2.7, tandis que la table 2.8 rapporte la moyenne des coefficients de régressions estimés pour une simulation avec et sans confusion, où les valeurs réelles sont entre parenthèses.

Rappelons que pour trouver un estimateur convergent de l'effet moyen du traitement, le modèle RD utilise des régressions locales où les individus sont comparables dans un voisinage de $[x_0 - h; x_0 + h]$ (8). Cette propriété est illustrée dans les résultats de l'impact de la présence d'une variable confondante, soit dans la table 2.7. Pour un $h = 30$, nous obtenons un biais estimé faible, mais un écart-type élevé. Pour $h = 183$, nous retrouvons notre compromis entre la variance et le biais qui agissent de façon contraire l'un avec l'autre ; un h élevé engendre un biais plus grand, mais un écart-type plus petit. Nous constatons aussi que les valeurs de l'ETM sont proches de celles de l'écart-type, ce qui indique que notre méthode d'estimation de la variabilité est adéquate. De plus, nous avons des probabilités de couverture dont les valeurs s'avèrent être relativement proches de la valeur nominale de confiance 95%.

Tableau 2.7: Impact de la confusion sur la largeur de bande pour une simulation Monte-Carlo avec 1000 réplications et une taille échantillonnale de 10 000, $\tau = -0.5913$

Largeur de bande	Moyenne	Biais	Écart-type	RCEQM	ETM	Probabilité de couverture
30	−0.613	−0.022	0.519	0.520	0.517	0.953
92	−0.587	0.005	0.288	0.288	0.293	0.945
183	−0.544	0.047	0.212	0.217	0.208	0.937

Pour fins de comparaison, nous avons refait une simulation sans confusion, mais avec une taille d'échantillon comparable de 10 000. Ainsi, la table 2.8 nous permet de constater que les coefficients estimés, dans le scénario avec confusion, sont erronés au niveau des ordonnées à l'origine, mais que leur différence $\alpha_D - \alpha_G$ demeure relativement proche de la vraie valeur du τ. En d'autres termes, il existe un biais des estimateurs des ordonnées à l'origine qui est plus important que celui de leur différence (voir table 2.7). Nous remarquons aussi que, dans le cas du scénario sans confusion, les coefficients sont estimés sans biais et que, par conséquent, on obtient des ordonnées à l'origine estimées adéquates.

De plus, les faibles biais estimés obtenus (voir table 2.7) suggèrent qu'en pratique la confusion est probablement négligeable dans un contexte réel de diagnostic de TDAH. En conclusion, les résultats démontrent la capacité du modèle de régression par discontinuité à fournir un estimateur non biaisé, dans une petite fenêtre de temps, en présence de confusion.

Tableau 2.8: Moyenne des coefficients estimés de la régression, pour une simulation Monte-Carlo avec 1000 réplications et une taille échantillonnale de 10 000, $\tau = -0.5913$

Largeur de bande	α_D (-3.178)	β_D (0.002)	β_G (0.002)	α_G (-2.587)
Scénario 1 : Avec confusion				
30	-3.483	0.002	0.003	-2.892
92	-3.484	0.002	0.002	-2.893
183	-3.520	0.002	0.001	-2.929
Scénario 2 : Sans confusion				
30	-3.209	0.002	0.001	-2.618
92	-3.194	0.002	0.001	-2.603
183	-3.188	0.002	0.002	-2.597

CHAPITRE III

APPROCHE BAYÉSIENNE

Dans ce chapitre, nous adoptons une perspective bayésienne pour vérifier quels sont les effets du choix des lois a priori sur l'estimation de l'effet moyen du traitement. Les résultats de cette investigation sont également comparés à ceux de l'approche fréquentiste afin de déterminer si l'un nous fournit des estimateurs plus performants que l'autre selon le critère de l'erreur quadratique moyenne (iii).

3.1 Objectifs de simulation

Rappelons que l'inférence bayésienne en RD requiert la spécification d'une loi a priori sur les paramètres inconnus de régression, comme dans le cas linéaire détaillé à la section 1.3.2. Ainsi, une des motivations qui encouragent l'utilisation de l'analyse bayésienne est la possibilité d'incorporer nos présuppositions dans le choix des lois a priori des paramètres d'intérêt. Ceci revient à imposer des contraintes significatives sur les valeurs de ces paramètres.

Une des méthodes d'échantillonnage a posteriori couramment utilisées est la méthode Monte-Carlo par chaînes de Markov (MCMC : "Markov-Chain Monte Carlo") qui, comparativement à l'approche fréquentiste, permet une plus grande flexibilité dans la structure de modélisation et fournit les estimations de mesures ciblées de façon simple et directe (Geneletti *et al.*, 2015). Dans son article d'ana-

lyse bayésienne de la RD, (Geneletti *et al.*, 2015), qui se concentre sur le cas du modèle RD "fuzzy", souligne l'importance d'une réflexion approfondie sur la spécification a priori. Ces auteurs suggèrent notamment qu'il n'est pas difficile, dans certaines situations, de formuler des présuppositions a priori plausibles et réalistes. La présente étude vise similairement à utiliser la simulation bayésienne pour vérifier l'impact d'un choix de loi a priori approprié (ou non) sur l'exactitude de nos estimateurs.

Dans le contexte de TDAH, il existe une littérature pouvant informer sur le comportement du paramètre d'intérêt ce qui n'est pas nécessairement le cas dans un autre contexte d'application de la RD. Il est donc dans l'ordre d'inclure ces informations dans la spécification de nos lois a priori. Par contre, le choix de celles-ci ne devrait pas être pris avec l'intention d'influencer l'inférence, afin qu'elle concorde avec nos conclusions prédéterminées par rapport au paramètre d'intérêt (Gelman *et al.*, 2004). Dans le cas où il y a peu d'informations ou d'hypothèses disponibles, l'utilisation de lois a priori plus diffuses est recommandée (Geneletti *et al.*, 2015).

Comme nous l'avons remarqué dans les sections précédentes, un plus petit h entraîne une augmentation de la variabilité des estimateurs et, en contrepartie, diminue le potentiel de biais de ceux-ci. Un de nos objectifs est de vérifier si, selon les différentes largeurs de bande, les estimateurs bayésiens ont une moins grande erreur que leurs contreparties fréquentistes.

3.1.1 Notions préliminaires

Les méthodes de Monte-Carlo par chaînes de Markov fournissent, à l'aide de simulations aléatoires, une caractérisation de la loi a posteriori des paramètres inconnus sans avoir l'obligation de connaître parfaitement les propriétés de celle-ci. Pour bien comprendre le rôle que joue chaque composante dans le processus de simula-

tion bayésien, nous résumons quelques propriétés des chaînes de Markov et de la méthode Monte-Carlo.

Soit une suite de variables aléatoires $\{X_t\}_{t\geq 0} = \{X_0, X_1, \dots\}$, un processus stochastique, alors on appelle chaîne de Markov à temps discret tout processus stochastique possèdant la propriété markovienne suivante :

$$P(X_t = i | X_{t-1} = j, X_{t-2}, \dots, X_0) = P(X_t = i | X_{t-1} = j),$$

où $i, j \in S$, l'espace des états discrets (Potvin, 2019). Rappelons que la définition de la propriété markovienne s'applique, de façon analogue, au cas d'espace d'état continu. Aussi, la probabilité de se déplacer d'un état i au temps t à un état j au temps $t+1$ est appelée probabilité de transition et est définie par la matrice $\mathbf{P} = p_{i,j} = P(X_{t+1} = j | X_t = i)$.

En d'autres mots, la probabilité conditionnelle de l'état futur X_{t+1} dépend uniquement de l'état d'aujourd'hui X_t ; le processus stochastique est alors dit sans mémoire.

Nous poursuivons en énumérant les propriétés principales (Mazet, 2003) des chaînes de Markov pour les méthodes de simulation MCMC :

1. Stationnarité : Soit π, la loi de la chaîne de Markov, alors π est stationnaire si, d'un temps à l'autre, elle demeure inchangée. Dans le cas discret, π est définie à l'état j par :

   $$\pi_j = \sum_i \pi_i p_{i,j},$$

   qui est équivalent en écriture matricielle à :

   $$\pi = \pi \mathbf{P}.$$

2. Irréductibilité : Une classe est un sous-ensemble de tous les états qui communiquent entre eux. Une chaîne de Markov est dite irréductible si elle ne comporte qu'une seule classe.

3. Récurrence : Lorsqu'un processus stochastique après avoir accédé à l'état i y revient avec probabilité 1, alors i est un état récurrent. Lorsque cela s'applique pour tous les états de la chaîne, alors celle-ci est récurrente.

4. Apériodicité : L'état i est apériodique si sa période, d, est égale à 1. C'est-à-dire qu'il existe deux moments consécutifs, t et $t+1$, où $p_{ii}^{(t)} > 0$ et $p_{ii}^{(t+1)} > 0$. Si tous les états de la chaîne sont apériodiques, la chaîne est apériodique.

5. Réversibilité : Une chaîne de Markov est dite réversible si la probabilité d'être à l'état i et d'aller à l'état j est la même que d'être à l'état j et aller à l'état i :

$$\pi_i p_{ij} = \pi_j p_{ji}, \forall i, \forall j.$$

Rappelons que l'inférence bayésienne requiert la connaissance de la loi a posteriori des paramètres d'intérêt, or, il est souvent impossible ou ardu de trouver une forme explicite de celle-ci. Par conséquent, l'estimation de la loi a posteriori va passer par la construction d'une chaîne de Markov irréductible, récurrente, apériodique et réversible, soit une chaîne ergodique, qui admet une loi stationnaire unique coïncidant avec la loi a posteriori (Defossez, 2019) :

$$X_t \xrightarrow[t\to\infty]{\mathcal{L}} \pi.$$

Nous poursuivons avec la méthode Monte-Carlo qui se définit comme étant une technique d'échantillonnage aléatoire dont le but est de calculer des intégrales (Grenon-Godbout, 2015). Soit une fonction de densité, $f_X(x)$, définie par souci de simplicité, sur le support d'une loi uniforme $[a, b]$, alors l'espérance mathématique de la fonction $g(X)$ se calcule de la façon suivante :

$$G = \mathbb{E}(g(X)) = \int g(x) f_X(x) dx.$$

L'objectif est de générer un échantillon aléatoire $(X_1, X_2, \ldots, X_t)$ de variables i.i.d. de densité $f_X(x)$ sur le support $[a, b]$ et de calculer une estimation de G, dite de

Monte-Carlo, par la moyenne empirique :

$$\tilde{g}_t = \frac{1}{t}\sum_{i=1}^{t} g(x_i).$$

L'estimateur de Monte-Carlo détient les propriétés suivantes :

(1) Par la loi des grands nombres :

$$\tilde{g}_t \xrightarrow[t\to\infty]{} \mathbb{E}(g).$$

(2) Nous quantifions la précision de notre estimateur, selon l'équation suivante :

$$\begin{aligned}\mathbb{V}(\tilde{g}_t) &= \frac{1}{t}\mathbb{V}(g(x)) \\ &= \frac{1}{t}\int g^2(x) f_x(x) dx - G^2.\end{aligned}$$

En pratique $\mathbb{V}(\tilde{g}_t)$ n'est pas connue. Or, puisque l'échantillon est i.i.d., nous utilisons la variance empirique pour l'estimer :

$$\mathbb{S}_{\tilde{g}_t} = \frac{1}{t-1}\sum_{i=1}^{t}(g(x_i) - \tilde{g}_t)^2 \simeq \mathbb{V}(\tilde{g}_t).$$

(3) Par le théorème limite centrale :

$$\frac{\tilde{g}_t - G}{\sqrt{\mathbb{V}(\tilde{g}_t)/t}} \sim N(0,1), \quad t \to \infty.$$

Suivant les propriétés des chaînes de Markov précédemment énumérées, il s'en suit que pour toute valeur initiale X_0 la propriété (1) de la méthode Monte-Carlo est respectée :

$$\tilde{g}_t \xrightarrow[t\to\infty]{} \mathbb{E}_\pi(g).$$

Notons que, contrairement au contexte traditionnel de Monte-Carlo, puisque les observations X_t ne sont pas indépendantes, l'expression de la variance, illustrée à la propriété (2), n'est pas applicable dans un MCMC.

Nous rappelons que le but est de construire une chaîne de Markov qui possède une loi stationnaire, π, telle que $\pi = \mathbb{P}$ où $\mathbb{P}$ représente notre loi a posteriori. Pour ce faire, il existe un algorithme simple à implémenter, le Metropolis-Hastings (M-H)

(Metropolis *et al.*, 1953), qui, comme le propose (Betancourt, 2017), se définit en deux étapes : la proposition et la correction. La première se réfère à toutes les perturbations stochastiques de l'état inital, tandis que l'autre rejette toutes propositions qui s'éloignent trop de la loi d'intérêt (posteriori). Plus précisément, nous avons l'algorithme suivant :

1. Initialisation : Choisir un point de départ X_0.
2. Pour $t = 1, 2, \ldots, T$, où T est un nombre entier non-négatif :
   (a) Proposition : On échantillonne un état (proposition) Z_t de la loi $\mathbb{D}(.|X_{t-1})$, tel que :
   $$Z_t \sim \mathbb{D}(.|X_{t-1}).$$
   (b) Correction : La probabilité d'accepter l'état proposé est donnée par :
   $$\omega(X_{t-1}, Z_t) = min\left(1, \frac{\mathbb{P}(Z_t)\mathbb{D}(X_{t-1}|Z_t)}{\mathbb{P}(X_{t-1})\mathbb{D}(Z_t|X_{t-1})}\right).$$
   Si l'on accepte alors $X_t = Z_t$, sinon $X_t = X_{t-1}$.

3.2 Spécification du modèle

Notre scénario de simulation se réfère à notre modèle de régression initial, soit celui sans l'ajout de variable explicative, ni de confusion.

Rappelons que notre première section (1.3.2) sur l'approche bayésienne se référait au cas commun de la réponse continue, or, dans notre contexte RD avec réponse binaire, le processus est différent. En particulier, nous devons prendre en compte seulement le choix des lois a priori pour les coefficients de régression, sans se préoccuper du lien avec la variance. Ainsi, nous réutilisons l'équation de référence suivante quant au modèle de régression logistique :

$$logit\Big(P(Y_i = 1|X_i = x, T_i = t)\Big) = \alpha_G + \beta_G(X_i - x_0) + \tau T_i + (\beta_D - \beta_G)(X_i - x_0)T_i.$$

Nous pouvons donc choisir des lois a priori normales indépendantes pour nos paramètres $(\alpha_G, \beta_G, \alpha_D, \beta_G)$ telles que :

$$\alpha_j \sim N(\mu_{\alpha,j}, \sigma^2_{\alpha,j}) \quad \text{et} \quad \beta_j \sim N(\mu_{\beta,j}, \sigma^2_{\beta,j}) \quad j = G, D.$$

Notons que la loi a priori sur nos ordonnées à l'origine se traduit par une loi a priori de la différence τ, tel que :

$$\alpha_D - \alpha_G \sim N(\mu_{\alpha,D} - \mu_{\alpha,G}, \sigma^2{}_{\alpha_D} + \sigma^2{}_{\alpha_G}).$$

Il s'en suit que nous spécifions les lois normales et centrons, pour la loi a priori informative, les moyennes a priori selon les valeurs précédemment utilisées pour générer notre modèle de simulation initial (voir section 2.2). La table 3.1 qui suit regroupe quatre différentes lois a priori qui varient selon leur niveau d'information. Nous utiliserons cette information lors de l'interprétation subséquente des résultats.

Tableau 3.1: Choix des hyperparamètres pour les lois a priori normales des coefficients de régression selon leur niveau d'information

Coefficients	μ	σ
Informative		
α_D	-3.178	0.5
α_G	-2.587	0.5
β_D	0	10
β_G	0	10
Non informative		
α_D	0	100
α_G	0	100
β_D	0	100
β_G	0	100
Erronée faible		
α_D	2.5	2.5
α_G	1	2.5
β_D	0	10
β_G	0	10
Erronée forte		
α_D	2.5	0.05
α_G	1	0.05
β_D	0	10
β_G	0	10

Ainsi, nous retrouvons les lois a priori suivantes pour notre paramètre d'intérêt :

$$\tau_{Inf} \sim N(-0.591, 0.707^2)$$
$$\tau_{Ninf} \sim N(0, 141.42^2)$$
$$\tau_{Errfaib} \sim N(1.5, 3.535^2)$$
$$\tau_{Errfor} \sim N(1.5, 0.071^2),$$

où Inf, $Ninf$, $Errfaib$ et $Errfor$ représentent respectivement la loi a priori informative, non informative, erronée faible et erronée forte.

3.3 Méthodologie et convergence

3.3.1 Introduction au Monte-Carlo Hamiltonien

Il existe plusieurs bibliothèques en R (R Core Team, 2017) pour traiter l'estimation bayésienne (BUGS (Lunn *et al.*, 2000a), JAGS (Plummer, 2003), etc.))[1]. Dans le cas de notre étude de simulation, nous utilisons la bibliothèque rstan (Stan Development Team, 2019) qui fournit un interface pour interagir avec le logiciel d'estimation bayésienne Stan (Stan Development Team, 2018). À travers la plateforme de R, Stan utilise une branche du MCMC pour tirer des échantillons aléatoires d'une loi a posteriori, soit le Monte-Carlo Hamiltonien (MCH).

Ce dernier est une méthode basée sur la physique (dynamique hamiltonienne) qui, à l'aide d'un processus stochastique, fournit une chaîne de Markov ergodique capable de produire de larges transitions avec une grande probabilité d'acceptation (Girolami et Calderhead, 2011). Ainsi, l'échantillonnage MCH introduit un vecteur de variables auxiliaires indépendantes, $s_t \in R^D$, tel que notre espace de paramètres

1. Voir Annexe E pour une description des logiciels implémentant les modèles de simulations bayésiennes.

qui était à D-dimensions passe, désormais, à $2D$-dimensions (Betancourt, 2017).

Nous pouvons redéfinir la loi visée comme étant une loi conjointe $\mathbb{T}(p_t, s_t)$. En supposant les lois indépendantes, celle-ci se décompose en deux termes selon la factorisation suivante :

$$\mathbb{T}(p_t, s_t) = \mathbb{P}(p_t)\mathbb{S}(s_t),$$

où $\mathbb{P}$ est la loi a posteriori du vecteur de variables aléatoires, $p_t \in R^D$, et $\mathbb{S}$ est la loi du vecteur de variables auxiliaires.

Nous pouvons donc écrire la loi conjointe selon la fonction Hamiltonienne $H(p_t, s_t)$:

$$\mathbb{T}(p_t, s_t) = \exp(-H(p_t, s_t)).$$

Nous poursuivons en prenant le logarithme négatif de la loi conjointe, tel que :

$$\begin{aligned} H(p_t, s_t) &= -log(\mathbb{T}(p_t, s_t)) \\ &= -log(\mathbb{S}) - log(\mathbb{P}) \\ &= K(s_t) + V(p_t), \end{aligned}$$

où $K(s_t)$ représente l'énergie cinétique et $V(p_t)$, qui correspond à la loi de la distribution ciblée (loi a posteriori), représente la somme d'énergie potentielle. Nous avons donc le système d'équations dynamique suivant :

$$\begin{aligned} \frac{dp_t}{dt} &= \frac{\partial H}{\partial s_t} = \frac{\partial K}{\partial s_t} \\ \frac{ds_t}{dt} &= -\frac{\partial H}{\partial p_t} = -\frac{\partial K}{\partial p_t} - \frac{\partial V}{\partial p_t}. \end{aligned}$$

L'objectif est de résoudre ce système en utilisant une classe d'intégrateurs numériques, plus spécifiquement l'intégrateur "leapfrog", ce qui permet de proposer de nouvelles réalisations pour le vecteur de paramètres et la composante cinétique.

Précédemment, la probabilité d'acceptation-rejet de l'algorithme MH (voir 2b) prenait en compte la loi visée, $\mathbb{P}$, or, la loi conjointe transforme la probabilité de décision qui doit, désormais, aussi comprendre la loi de s_t. Ainsi, nous acceptons le nouvel état (p_t^*, s_t^*) avec la probabilité mise à jour :

$$min\left(1, exp\{-H(p_t^*, s_t^*) + H(p_t, s_t)\}\right),$$

sinon on rejette et garde (p_t, s_t). Ces nouvelles trajectoires proposent des états qui sont éloignés du point courant, ce qui permet d'explorer efficacement l'espace des états ; ces propositions, dites intelligentes, ont une plus grande probabilité d'acceptation que dans le cas des MCMC traditionnels (Girolami et Calderhead, 2011).

Il s'en suit, qu'en tenant compte de la réversibilité du système d'équation, la loi jointe et, par conséquent, les lois marginales, $\mathbb{S}(s_t)$ et $\mathbb{P}(p_t)$, sont inchangées ((Girolami et Calderhead, 2011),(Defossez, 2019)) ; nous pouvons donc retrouver la loi a posteriori ciblée de façon marginale.

3.3.2 Diagnostic de convergence

La bibliothèque **rstan** détient plusieurs options qui, lorsque fixées de façon judicieuses, permettent une implémentation optimale de notre simulation bayésienne. Une de celles-ci est le nombre de chaînes MCMC qui produisent des échantillons indépendants d'une chaîne à l'autre.

Chaque chaîne part d'un point initial différent et typiquement à l'écart des autres points ; ainsi, en lançant plusieurs chaînes on vérifie la convergence du processus. Dans le but de faciliter le roulement d'une simulation complexe, il est possible de sélectionner un nombre de noyaux ("core"). Par exemple, pour quatre chaînes, le choix adéquat serait de quatre noyaux, car cela permettrait de rouler quatre chaînes indépendantes en parallèle (Ovando, 2018). Le nombre d'itérations par

chaîne est aussi une option qu'il est nécessaire de spécifier. Par défaut, la première moitié des valeurs échantillonnées est traitée comme des échantillons tests et est rejetée ("warmup").

Dans notre cas, puisque nous nous intéressons à un modèle de simulation assez simple, il est suffisant de prendre une seule chaîne avec 10 000 itérations. Pour s'assurer que nos choix sont adéquats, nous présentons le graphique de diagnostic de convergence ("traceplot") à la figure 3.1 qui vérifie la fiabilité de notre simulation en s'assurant qu'il n'y a pas de problèmes apparents quant à la convergence vers notre loi a posteriori. Notons que la vérification de convergence s'est faite à l'aide de la loi a priori informative.

Ce graphique illustre, pour notre paramètre d'intérêt, τ, les valeurs de l'échantillon MCMC après chaque itération successive dans la chaîne. Ainsi, tout schéma est une indication d'un problème de convergence et qu'un plus grand nombre d'itérations est requis pour garantir que la loi a posteriori est correctement représentée (Logan, 2018). La figure 3.1 nous suggère que la décision d'écarter les 1000 premières simulations est appropriée pour notre simulation.

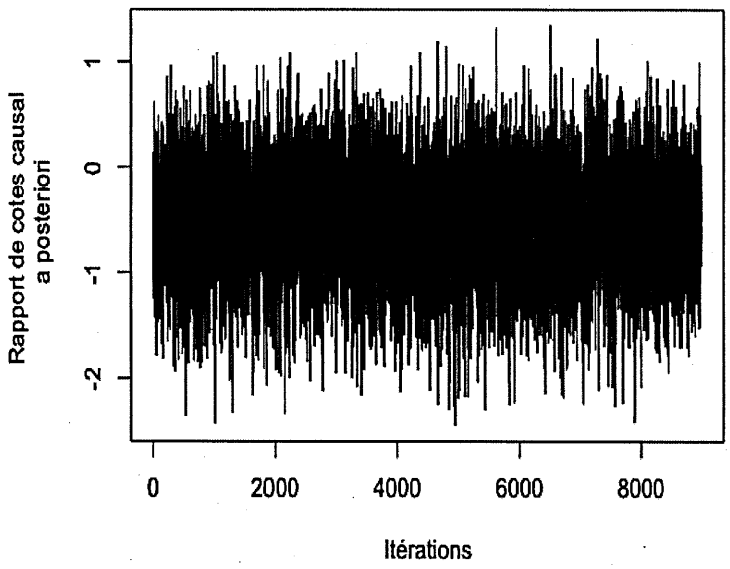

Figure 3.1: Diagnostic des échantillons tests

3.4 Résultats et analyses a posteriori

Nous tentons, au cours de cette section, de tester le caractère robuste de notre algorithme de simulation bayésien lors d'un mauvais choix de loi a priori. De plus, nous nous intéressons à la performance des estimateurs bayésiens par rapport à ceux du cas fréquentiste. Nous avons donc les cinq scénarios suivants : une loi a priori informative, une loi a priori non informative, une loi a priori erronée faible, une loi informative erronée forte et les résutats de notre étude de simulation fréquentiste.

La table 3.2 nous rapporte les résultats de cette comparaison. Rappelons que les choix pour les quatre types de loi a priori sont détaillés à la table 3.1 et que les

données du cas fréquentiste proviennent de la table 2.2. Notons également que nous avons calculé les estimateurs bayésiens à l'aide des mêmes données simulées de la section 2.2.

En premier lieu, nous remarquons, pour la loi a priori informative, qu'il y a une différence importante entre les résultats de l'écart-type estimé pour $h = 30$ et $h = 183$. En effet, la première est respectivement plus petite et l'autre plus grande par rapport au cas fréquentiste, tandis que le biais estimé reste légèrement plus élevé pour la loi a priori informative. Il s'en suit que la plus petite largeur de bande coïncide avec une meilleure perfomance de la simulation bayésienne. En général, nous avons des biais qui varient selon la fenêtre de temps, mais demeurent relativement faibles, par conséquent, la RCEQM décroît au même ordre que l'écart-type.

Du côté de la loi non informative, nous constatons que ses résultats se rapprochent le plus de ceux du cas fréquentiste. Comme dans le scénario précédent, la plus petite largeur de bande $h = 30$ correspond à un écart-type moins élevé que celui fréquentiste, mais le biais estimé et la RCEQM sont plus grands. Pour ce qui est des deux autres largeurs de bande, soient $h = 92$ et $h = 183$, les écarts-type de la loi a priori non informative sont tous plus élevés que ceux fréquentistes. Finalement, nous remarquons que les biais estimés sont plus importants que ceux du cas fréquentiste, mais qu'en général leurs valeurs restent atténuées ; les variations de la RCEQM concordent avec celles des écart-types.

Le choix des hyperparamètres des deux lois a priori erronée s'est fait tel que la vraie valeur du τ se situe dans la queue de leur distribution respective. Par conséquent, nous nous attendons à ce que ces lois engendrent des résultats moins adéquats que les autres, ce qui est démontré par les valeurs trouvées. En effet, parmi tous les scénarios, les biais estimés de la loi a priori erronée forte sont beaucoup plus élevés que ceux du cas fréquentiste dû à une estimation totalement incorrecte du

τ. Nous retrouvons la même situation pour la loi a priori erronée faible, mais dans une optique plus modérée à cause d'une surestimation moins importante. Pour la loi a priori erronée forte, les valeurs de la RCEQM coïncident avec celles des biais estimés, tandis que, pour la loi a priori erronée faible, le comportement de la RCEQM concorde avec celui des écart-types. Rappelons, que nous avons spécifié, lors de la sélection des hyperparamètres de la loi a priori erronée forte, une plus petite variance pour les paramètres d'ordonnées à l'origine. Il s'en suit que les écarts-types, pour toutes largeurs de bande, sont très petits.

Pour conclure, la loi a priori informative ainsi que celle non informative ont généralement bien performé ; les résultats, pour la loi a priori informative, ont particulièrement surpassés ceux fréquentistes pour la plus petite largeur de bande $h = 30$. Il demeure que le scénario le moins performant est celui de la loi a priori erronée forte, suivie par celle faible, et ce, pour toutes largeurs de bande.

Tableau 3.2: Comparaison de l'estimation de l'effet du traitement selon la méthode fréquentiste (scénario 5) et une simulation bayésienne selon différents choix de lois a priori (scénarios 1-4) pour une simulation Monte-Carlo de 1000 et une taille d'échantillon de 500, $\tau = -0.5913$

Largeur de bande	**Moyenne**	**Biais**	**Écart-type**	**RCEQM**
Scénario 1 : Loi a priori informative				
30	−0.564	0.028	0.651	0.652
92	−0.598	−0.007	0.574	0.574
183	−0.633	−0.042	0.506	0.508
Scénario 2 : Loi a priori non informative				
30	−0.645	−0.054	1.126	1.298
92	−0.619	−0.028	0.580	0.580
183	−0.601	−0.010	0.403	0.403
Scénario 3 : Loi a priori erronée faible				
30	−0.736	−0.145	1.961	1.966
92	−0.690	−0.099	1.172	1.176
183	−0.697	−0.106	0.833	0.840
Scénario 4 : Loi a priori erronée forte				
30	1.475	2.067	0.070	2.068
92	1.420	2.011	0.070	2.012
183	1.334	1.925	0.070	1.926
Scénario 5 : Résultats fréquentistes				
30	−0.594	−0.002	1.178	1.178
92	−0.581	−0.010	0.569	0.569
183	−0.595	−0.003	0.396	0.396

3.4.1 Analyse a posteriori prédictive

Rappelons que l'un des avantages de la simulation bayésienne est la possibilité d'ajuster notre modèle en fonction de présuppositions guidant le choix des lois a priori des coefficients de régression. Par le fait même, si nos hypothèses a priori sont réalistes, alors notre modèle devrait être en mesure de générer des données qui se rapprochent le plus possible de celles observées ; ainsi l'analyse a posteriori prédictive nous permet d'évaluer l'ajustement de notre modèle.

Considérons le vecteur de nos paramètres d'intérêt $\theta = (\alpha_G, \beta_G, \alpha_D, \beta_G)$, alors la loi a posteriori prédictive peut être écrite de la façon suivante :

$$p(\widetilde{Y}|Y,X) = \int p(\widetilde{Y}|\theta,Y,X)p(\theta|Y,X)d\theta,$$

où $\widetilde{Y}$ est le vecteur de la réponse prédite non observée, Y est le vecteur de la réponse observée et X est celui des variables explicatives.

Or, en supposant que $\widetilde{Y}$ est indépendant de Y conditionnellement à θ, l'équation précédente se simplifie :

$$p(\widetilde{Y}|Y,X) = \int p(\widetilde{Y}|\theta,X)p(\theta|Y,X)d\theta,$$

Ainsi, pour chaque tirage $s = 1, \ldots, S$ du vecteur de paramètres par la loi a posteriori, $\theta^{(s)} \sim p(\theta|Y,X)$, on échantillonne un vecteur $\widetilde{Y}^{(s)}$ de taille n du modèle logistique $p(\widetilde{Y}|\theta^{(s)},X)$; il en résulte un échantillon de n réponses prédites provenant de la loi a posteriori prédictive $p(\widetilde{Y}|Y,X)$ (Gabry, 2019). Il s'en suit que si nos hypothèses sont réalistes, alors ce processus de simulation va générer des données similaires à celles que l'on a actuellement observées.

La figure 3.2 nous présente les résultats de cette analyse prédictive a posteriori, pour un jeux de données, où l'on illustre les probabilités empiriques d'être diagnostiqué TDAH (gris) versus non diagnostiqué (noir) pour chaque loi, ainsi que

pour la réponse observée. Nous remarquons immédiatement que la loi prédictive a posteriori erronée forte est celle qui concorde le moins avec les réponses observées. Par conséquent, le choix des hyperparamètres de cette loi n'a pas réussi à engendrer un modèle a posteriori prédictif compatible aux réponses observées. Celle-ci est suivie par la loi prédictive erronée faible et non informative ; la proportion de ces deux lois, qui se surperposent presque parfaitement, est plus grande que celle de la réponse observée, mais proche de la loi informative.

La figure 3.2 nous démontre que nous avons relativement fait les bonnes hypothèses quant aux choix des hyperparmètres pour la loi a priori informative. En effet, nous constatons que la proportion prédictive a posteriori d'être diagnostiqué TDAH a presque réussit à reproduire celle de la réponse observée ; parmi tous les scénarios, elle est celle dont la proportion se rapproche le plus de la proportion observée.

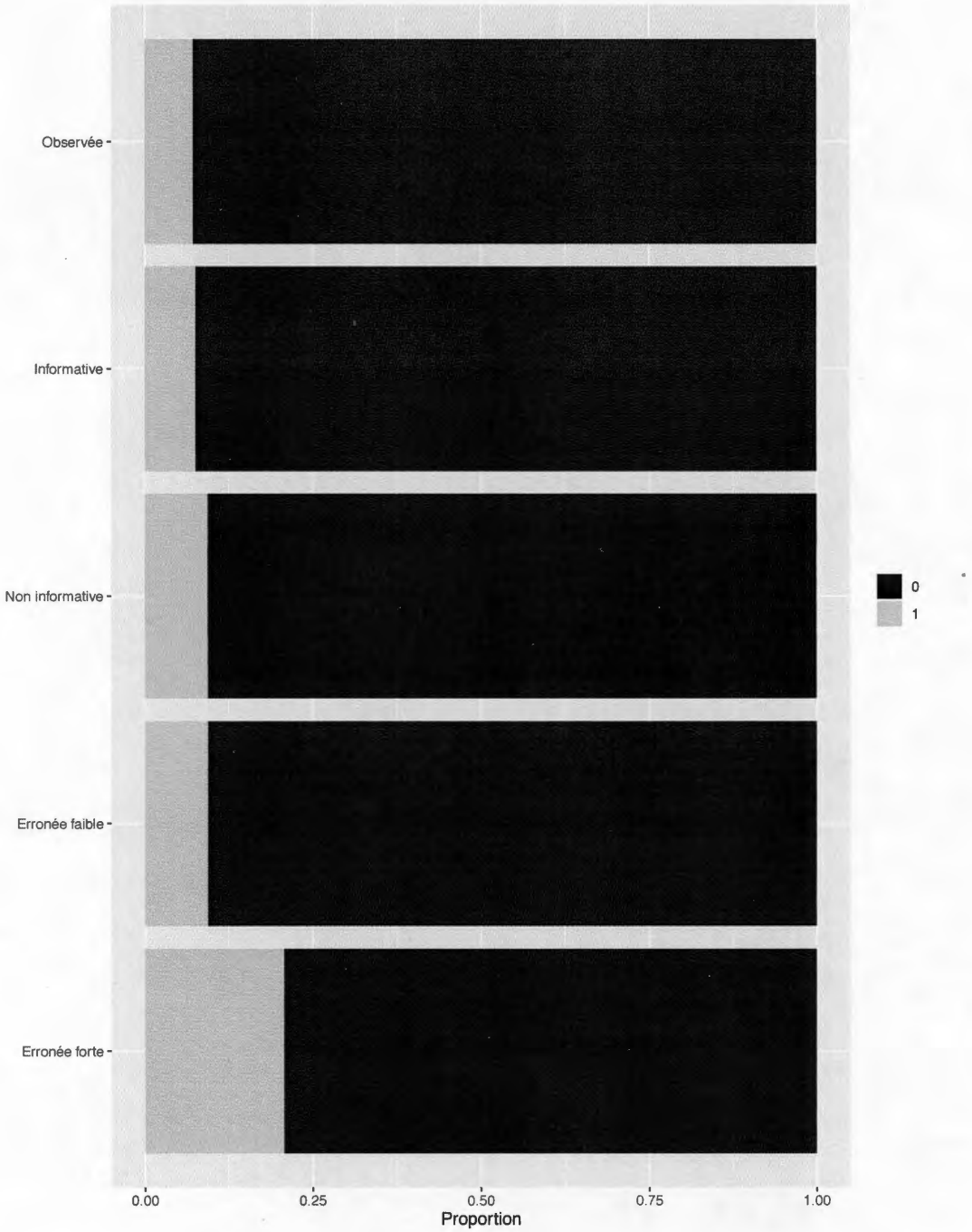

Figure 3.2: Comparaison de la proportion d'individus diagnostiqué TDAH selon la réponse observée et les différents choix de lois prédictives a posteriori, pour un jeux de données de 500 individus

CONCLUSION

Au cours de ce mémoire, nous avons étudié, à travers différents modèles de simulation, la régression par discontinuité pour une variable binaire. Ainsi, nos avons utilisé une mise en contexte spécifique lors de nos processus de simulations et d'analyse des résultats, soit l'association entre la probabilité d'être diagnostiqué TDAH et l'âge d'entrée des enfants à l'école.

Nous avons débuté par décrire le modèle contrefactuel et introduire brièvement le concept de confusion. Puis nous avons continué avec le modèle de régression par discontinuité "sharp" où nous avons expliqué les hypothèses définissant le processus d'application du modèle. Cette première section s'est terminée en présentant deux approches pour estimer l'effet moyen du traitement : la régression linéaire simple (fréquentiste) et sa contrepartie bayésienne.

Par la suite, nous avons présenté deux mesures pour estimer l'effet du traitement moyen : le rapport de cotes causal, lors de la régression logistique, et le risque relatif causal lors de la régression log-binomiale. Il est a noter que le rapport de cotes causal peut être vu comme une bonne approximation du risque relatif causal. En effet, puisque le cas du diagnostique de TDAH a une prévalence rare, nous avons pu interpréter l'effet du traitement moyen du modèle logistique comme un risque relatif causal. Il serait intéressant de valider cette hypothèse, en reprenant les analyses, dans le cadre d'un modèle log-binomial, et en considérant le risque relatif causal comme mesure d'estimation.

La section suivante marque le début de notre étude de simulation fréquentiste, où nous avons commencé par décrire l'algorithme standard de simulation pour, par

la suite, étudier différents scénarios de simulation. Ainsi, nous avons pu analyser l'impact de la taille d'échantillon et de la fenêtre de temps sur l'estimation de l'effet du traitement moyen ; la précision de l'estimateur augmente avec la taille d'échantillon et le biais n'est pas affecté par la diminution de la fenêtre de temps lorsque l'algorithme de simulation ne comprend pas de variable de confusion.

Nous avons ensuite examiné, dans un contexte précis, l'impact de l'ajout d'une variable explicative sur la précision des estimateurs. Nous avons observé que plus l'effet de la variable est important, plus on aurait avantage à ajuster cette variable pour obtenir, au prix d'une perte de précision, un gain au niveau de la puissance. De son côté, l'ajout de la variable de confusion nous a permis de retrouver la caractéristique principale d'un modèle de régression par discontinuité, soit l'estimation non biaisée dans un voisinage du point de discontinuité. Ainsi, nous avons obtenu, pour la plus petite fenêtre de temps, un biais estimé faible, mais un écart-type élevé.

La dernière section de ce mémoire comprend l'analyse de l'impact du niveau d'information des lois a priori sur l'estimation de l'effet du traitement moyen, lors d'un processus de simulation bayésienne. Pour ce faire, nous avons utilisé quatre types de loi a priori : informative, non informative, erronée faible et forte. Ainsi, nous avons observé que si la loi a priori est informative et bien spécifiée, nous obtenons relativement une bonne estimation pour une petite fenêtre de temps. Par la suite, nous avons retrouvé, que pour la loi a priori non informative, approximativement les résultats fréquentiste, tandis qu'une mauvais spécification de la loi a priori, telle que pour celle erronée forte et faible, se résulte en un impact négatif, surtout quand la fenêtre de temps est plus petite.

Tout au long de ce mémoire, nous avons supposé le respect parfait de la règle de décision, or, en pratique on s'attend plutôt que certains individus ne respectent pas

la règle d'affectation. Il serait donc intéressant d'analyser l'impact du non-respect de cette hypothèse sur l'estimation de l'effet du traitement. De plus, (N Evans *et al.*, 2010) rapportent, dans leur article, le rôle du parent et du professeur dans le processus de diagnostic du TDAH chez l'enfant. Analyser le rôle de cette information augmenterait la complexité du modèle de simulation et permettrait d'illustrer son impact sur celui-ci.

ANNEXE A

CALCUL DE LA PENTE

Pour calculer la pente, en terme des coefficients associé à notre variable d'affectation X_i^c, nous travaillons sur le logit de la probabilité directement.

1. Différence des ordonnées à l'origine :

$$\begin{aligned} dp &= log((1-0.04)/0.04) - log((1-0.07)/0.07) \\ &= log(0.96/0.04) - log(0.93/0.07) \\ &= 0.5913645 \end{aligned}$$

2. Nous voulons avoir une pente unitaire à mi-chemin entre les ordonnées à l'origine, tel que :

$$\begin{aligned} dpp &= dp/(2*183) \\ &= 0.001616 \end{aligned}$$

ANNEXE B

ÉQUIVALENCE DE LARGEURS DE BANDE ET TAILLES ÉCHANTILLONNALES

Le tableau suivant illustre les résultats pour les largeurs de bande équivalentes aux tailles d'échantillon respectives de 250, 500 et 1000. Lorsque nous comparons les résultats annalogues dans le cas de l'impact de la taille d'échantillon (section 2.2.1), nous obtenons des mesures pratiquement identiques. Dans le but de faciliter la comparaison, la table a été reproduite ci-dessous, où les valeurs en paranthèses correspondent aux résultats pour les tailles d'échantillons initiales.

Tableau B.1: Impact de la largeur de bande pour une simulation Monte-Carlo de 1000 et une taille échantillonnale de 1000, $\tau = -0.5913$

Largeur de bande	**Moyenne**	**Biais**	**Écart-type**	**RCEQM**	**ETM**
45	−0.605	−0.014	0.600	0.600	0.577
(250)	(−0.603)	(−0.011)	(0.580)	(0.581)	(0.575)
88	−0.599	−0.008	0.414	0.414	0.408
(500)	(−0.595)	(−0.003)	(0.396)	(0.396)	(0.401)
183	−0.584	0.008	0.280	0.280	0.282
(1000)	(−0.584)	(−0.008)	(0.280)	(0.280)	(0.282)

ANNEXE C

ILLUSTRATION DE LA DISTRIBUTION DES DATES DE NAISSANCE EN FONCTION DU STATUT SOCIO-ÉCONOMIQUE

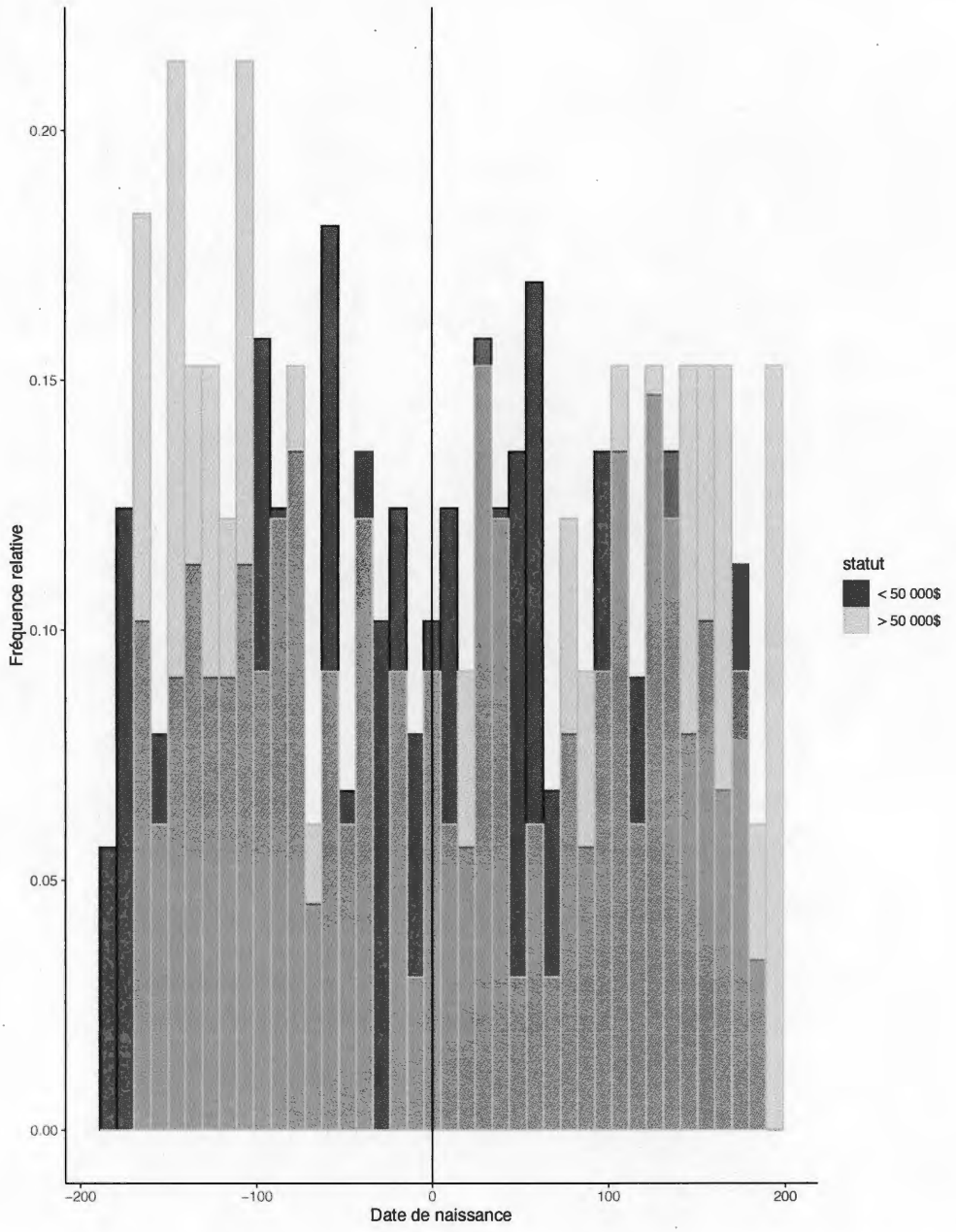

Figure C.1: Distribution empirique des dates de naissances ($X_i{}^c$) en fonction du statut économique

ANNEXE D

ILLUSTRATION DE LA DISTRIBUTION DU DIAGNOSTIC DE TDAH EN FONCTION DU STATUT SOCIO-ÉCONOMIQUE

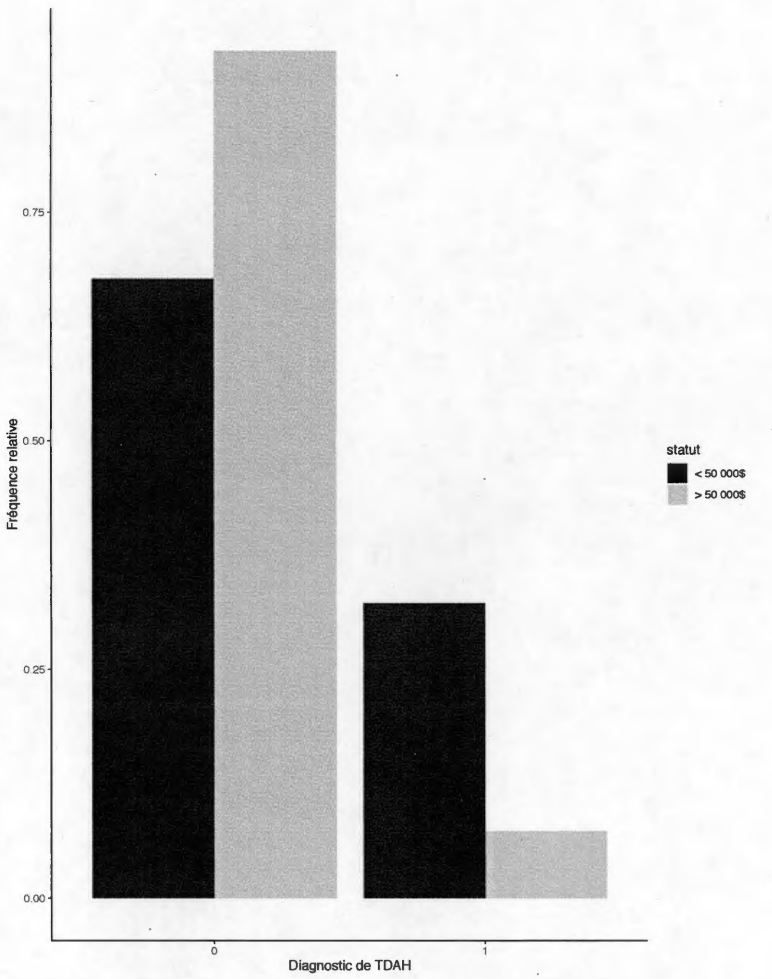

Figure D.1: Distribution empirique du diagnostic TDAH (Y_i) en fonction du statut socioéconomique

ANNEXE E

LOGICIELS DE SIMULATION

Nous présentons brièvement trois logiciels de simulation bayésienne : WinBUGS, JAGS et Stan. À noter que notre description va traiter de l'utilisation de ces logiciels à travers la plateforme de programmation R. Ceci possibilité offre aux utilisateurs une plus grande liberté dans la manipulation et la visualisation des données. En effet, il est possible de formater les données, générer des valeurs initiales, visualiser et sauvegarder les résultats de l'échantillonnage a posteriori avec de simples commandes. Notons également que nous ne rentrons pas dans les détails du fonctionnement des commandes de R ; nous supposons que les lecteurs sont familiers avec les notions de base.

Tout au long de cette section, nous allons utilisé un exemple de régression linéaire simple. Soit :

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i,$$

alors pour implémenter une forme bayésienne de régression linéaire, nous faisons l'hypothèse que la variable réponse (Y_i) suit une loi normale $N(\mu, \sigma^2)$ où μ est une fonction de nos deux paramètres β_1 et β_2. Ainsi, nous assignons les lois a priori suivantes :

$$\beta_1 \sim N(0, 100), \beta_2 \sim N(0, 100), \quad \text{et} \quad \sigma^2 \sim IG(0.001, 0.001).$$

E.1 BUGS

Le logiciel BUGS ("Bayesian inference Using Gibbs Sampling") est un système de programmation qui ajuste des modèles de simulation bayésiennes en utilisant l'échantillonnage de Gibbs. Ce dernier permet d'implémenter la méthode MCMC et d'estimer les distributions a posteriori des paramètres d'intérêts. Plusieurs versions ont suivi celle initiale, dont WinBUGS (Lunn *et al.*, 2000b) qui peut uniquement être utilisée sur la plateforme Windows et OpenBUGS (Surhone *et al.*, 2010) dont la flexibilité d'utilisation est plus importante que ses prédécesseurs (Depaoli *et al.*, 2016).

La bibliothèque R2WinBUGS (Sturtz *et al.*, 2005) permet aux utilisateurs d'interagir avec WinBUGS à travers R. Ainsi, on transfère, à l'aide de la fonction bugs, de WinBUGS à R les informations cruciales au fonctionnement du modèle de simulation. C'est avec ces informations que WinBUGS simule et échantillonne à partir des lois a posteriori des paramètres, pour ensuite retourner les échantillons a posteriori sur la plateforme R (Matzke *et al.*, 2017). À noter que les modèles de simulation doivent êtes écrits dans un fichier texte à part (ex : modele.txt) et que la vectorisation n'est pas supportée par WinBUGS, tel que l'on doit utiliser des boucles pour que notre modèle passe à travers toutes nos observations.

Les étapes suivantes décrivent le processus d'implémentation de la simulation bayésienne :

1. Installer et charger la biblothèque R2WinBUGS.
2. Spécification du modèle : La modélisation se fait en deux sections, soient une pour décrire la fonction de vraisemblance et une autre pour les lois a priori des paramètres.

```
#fonction de vraisemblance
```

```
model {
for(i in 1:N){
y[i] ~ dnorm(mu[i], tau)
mu[i] <- beta1 + beta2*x1[i]
}
# loi a priori
beta1 ~ dnorm(0, 100)
beta2 ~ dnorm(0,100)
sigma ~ dgamma(0.001,0.001)
}
```

3. Préparation de la base de données : Toutes les variables qui seront utilisées dans le modèle doivent être à l'intérieur d'une liste de nom qui inclut le nombre d'observations (N). Comme WinBUGS ne peut pas interpréter les tableau de R ("dataframe"), il est obligatoire de définir chaque variable dans un objet séparé. Soit la table df, alors :

```
X1 <- df$x1
N <- length(X1)
data <- list("N","X1")
```

4. Énumération des paramètres d'intérêt :

```
parameter <- c("beta1", "beta2","sigma")
```

5. Définition des valeurs initiales : WinBUGS peut définir, pour chaque paramètre, les valeurs initiales des chaînes. Il est aussi possible de les sélectionner soi-même à l'aide d'une fonction qui va rouler pour chaque chaîne.

```
inits <- function(){
list(beta1=rnorm(1), beta2=rnorm(1), sigma=runif(1,0,2))
}
```

6. Réglages MCMC :

```
sims <- bugs(model.file="modele.txt",
data = data,
parameters = parameter,
inits = inits,
n.chains = 2, # nombre de chaines MCMC
n.iter = 25000, # nombre iteration par chaine
n.burnin = 20000, # nombre echantillon test
n.thin = 5, # taux amincissemnt (thinning rate)
bugs.directory = "c:/Users/Francelyne/Documents/WinBUGS14/"
#emplacement ou WinBUGS est installe
)
```

Lorsque WinBUGS a terminé de rouler les simulations, il retourne les échantillons a posteriori vers R dans l'objet sims. Nous pouvons, par la suite, évaluer la convergence des chaînes MCMC, obtenir un graphique résumé des coefficients et des intervalles de crédibilité, sauvegarder les valeurs estimées pour chaque paramètre, etc. (Jarvis, 2015) et (Hare, 2014) offrent des tutoriels d'introduction simples et expliqués.

E.2 JAGS

JAGS ("Just Another Gibbs Sampler") a été créé dans le but d'être compatible avec BUGS, par conséquent, il utilise le même langage informatique que ce dernier. Par contre, son interface est plus flexible et sa plateforme d'accès n'est pas limitée à Windows (Depaoli *et al.*, 2016). Plusieurs bibliothèques de R fournissent une interface pour interagir avec le logiciel JAGS : rjags (Plummer, 2019), runjags (Denwood, 2016) et R2jags (Su et Yajima, 2015).

La bibliothèque rjags permet à l'utilisateur de facilement s'initialiser à l'implémentation d'un modèle bayésien avec JAGS. Tout d'abord, il faut, définir notre modèle de simulation en utilisant la fonction jags.model. Par la suite, nous pouvons nous servir de commandes pour extraire des échantillons, examiner la loi a posteriori et vérifier les diagnostics de convergence à l'aide d'autres bibliothèque comme coda (Depaoli *et al.*, 2016).

De leur côté, runjags et R2jags sont similaires à rjags, mais ont des options additionnelles permettant de modéliser des études de simulations plus complexes. Entre autres, runjags offre la possibilité de rouler de multiples chaînes en parallèle et d'automatiser la vérification des diagnostics de convergence (Depaoli *et al.*, 2016). R2jags offre aussi l'option des chaînes parallèles et, comme son interface d'utilisation est relativement similaire a R2WinBUGS, peut être utilisée pour convertir les données de WinBUGS au format de JAGS.

Nous nous servons de R2jags pour présenter l'implémentation d'un modèle de simulation bayésien à travers la plateforme R. Les étapes suivantes sont pratiquement identiques à celles énumérées pour R2WinBUGS, la seule différence étant la fonction utilisée pour échantillonner de la distribution a posteriori :

1. Installer et charger la biblothèque R2jags.
2. Spécification du modèle :

```
#fonction de vraisemblance
model {
for(i in 1:N){
y[i] ~ dnorm(mu[i], tau)
mu[i] <- beta1 + beta2*x1[i]
}
# loi a priori
```

```
beta1 ~ dnorm(0, 100)
beta2 ~ dnorm(0,100)
sigma ~ dgamma(0.001,0.001)
}
```

3. Préparation de la base de données :

```
X1 <- df$x1
N <- length(X1)
data <- list("N","X1")
```

4. Énumération des paramètres d'intérêt :

```
parameter <- c("beta1", "beta2","sigma")
```

5. Définition des valeurs initiales :

```
inits <- function(){
list(beta1=rnorm(1), beta2=rnorm(1), sigma=runif(1,0,2))
}
```

6. Réglages MCMC :

```
sims <- jags(data = data,
inits = inits,
parameters.to.save = parameter,
model.file="modele.txt",
n.chains = 2,
n.thin = 5,
n.iter = 25000,
n.burnin = 20000
)
```

Comme pour WinBUGS, lorsque JAGS met fin à son processus de simulation il retourne les échantillons a posteriori vers R dans l'objet sims. Il est donc possible,

par la suite, d'avoir accès aux résultats de les visualiser et d'en tirer un résumé. Pour un exemple complet d'application, veuillez-vous référer à (Kruschke, 2010).

E.3 Stan

Nous poursuivons avec Stan, l'unique logiciel qui opère en C++, ce qui lui confère une plus grande flexibilité dans la création de modèles plus complexes. À noter que Stan exige que chaque variable soit déclarée avant son utilisation et que son type numérique, ainsi que son domaine de définition, soient spécifiés.

La simulation par Stan est divisée en six blocs : données (data), données transformées (transformed data), paramètres (parameters), paramètres transformés (transformed parameters), modèle (model) et quantités générées (generated quantities). Tous les blocs, exculant "modèle", sont optionnels, mais l'ordre doit être préservé. Par exemple, il est possible d'avoir le bloc "données transformées " après celui "données" ; il contient les instructions pour transformer les objets du bloc "données" en de nouvelles valeurs pouvant être utilisées dans le modèle défini par la suite. De façon analogue, le bloc "paramètres transformés" contient les mêmes indications pour modifier les paramètres du bloc "paramètres". Finalement, le bloc "quantités générées" contient des mesures qui seront générées à partir du modèle à chaque étape telles que les valeurs prédictives ((Kruschke, 2010)).

Ainsi, comme mentionné à la section 3.3.1, la bibliothèque rstan permet aux utilisateurs de gérer Stan à travers l'interface de R. Les étapes suivantes permettent l'implémentation de simulation bayésienne à l'aide du logiciel :

1. Installer et charger la biblothèque rstan.
2. Spécification du modèle : À noter que, dans notre exemple, nous sauvegardons notre modèle Stan dans un vecteur caractère pour ensuite le passer

directement à la fonction d'échantillonnage de Stan. Par contre, il est possible d'obtenir des résultats identiques en sauvegardant le même modèle dans un fichier stan (mcmc.stan).

```
modele <-
<<data {
int<lower=0> N ;
// y est un vecteur de nombres entiers et de longueur N
int y[N] ;
int x1[N];

}
parameters {
vector[2] beta; // beta1, beta2
//specifie les limites pour la variance
real<lower=0,upper=1> sigma ;
}
model {
// definition des lois a priori
beta[1] ~ normal(0, 100);
beta[2] ~ normal (0,100);
// pour la variance
sigma ~ inv_gamma(0.001, 0.001);
// modele de regression lineaire
y ~ normal(beta[1] + beta[2] * x1 , sigma) ;
}
>>
```

3. Préparation de la base de données :

```
X1 <- df$x1
N <- length(X1)
```

```
data <- list("N","X1")
```

4. Énumération des paramètres d'intérêt :

```
parameter <- c("beta1", "beta2","sigma")
```

5. Définition des valeurs initiales :

```
inits <- function(){
list(beta1=rnorm(1), beta2=rnorm(1), sigma=runif(1,0,2))
}
```

6. Réglages MCMC :

```
sims <- stan(model_code = modele,
init=inits,
pars=parameter,
data = data,
iter = 25000,
chains = 2,
warmup = 20000,
thin = 5)
```

À la fin des simulations, Stan retourne les échantillons a posteriori vers R dans l'objet sims. Par la suite, il est possible d'accéder, de visualiser et de résumer les résultats des simulations bayésiennes. (Kruschke, 2010) offre une bonne documentation d'introduction, ainsi qu'un exemple complet d'application.

Pour un exemple comparatif des trois logiciels, veuillez-vous référer à l'article (Matzke *et al.*, 2017).

ANNEXE F

CODE DE SIMULATION EN R

```
###################################
# 1. Modele de simulation standard
###################################
## arguments:
# Xc2: variable d'affectation
# T_i : variable indicatrice du traitement
# y : reponse binaire
# p0 : probabilite de succes quand T_i=0
# p1 : probabilite de succes quand T_i=1
# alphaD = -3.178
# betaD = 0.00162
# alphaG = -2.5867
# betaG = 0.00162
###################################
## bibliotheques a utiliser
library(broom)
library(purrr)
library(dplyr)
library(tidyr)
```

```
# Fonction de simulation
SimulationData<-function(nTaille,alphaD, betaD, alphaG, betaG){
Xc2 <- round(runif(nTaille,-183,183))
T_i <- ifelse(Xc2>=0,1,0)
z0 <- alphaG + betaG*Xc2
p0 <- 1/(1+exp(-z0))
z1 <- alphaD + betaD*Xc2
p1 <- 1/(1+exp(-z1))
y1 <- rbinom(nTaille, 1, p1)
y0 <- rbinom(nTaille, 1, p0)
y <- ifelse(T_i==1,y1,y0)
df <- data.frame(Xc2, T_i, y)
#filtre quand la fenetre de temps change: tt <- df %>%
# filter(Xc2 %in% (-30:30))
fit <- glm(y~Xc2*T_i, family = "binomial", data=df)
}

# extraire les resultats de la simulation
simulation <- replicate(1000,SimulationData(500,-3.178,0.00162,
-2.5867,0.00162),simplify = FALSE)
names(simulation) <- paste0("SIMULATION",1:1000)
all_coefs <-plyr::ldply(simulation, tidy, .id = "simulation")

# rearrange les resultats
results <- all_coefs %>%
dplyr::select(-(statistic:p.value))  %>%
pivot_wider(
names_from = term,
values_from = c(estimate, std.error)) %>%
```

```
select(-c("std.error_(Intercept)","std.error_Xc2","std.error_Xc2:T_i"))

################################################
# 2. Fonction de simulation pour la confusion
################################################
## argument :
# statut : variable de confusion
################################################

SimulationData<-function(nTaille,alphaD, betaD, alphaG, betaG){

statut <- sample( 0:1, nTaille, replace=TRUE, prob=c(0.7139,0.2862))
Xc2 <- round(runif(nTaille,-183,183) + (30*statut))
T_i <- ifelse(Xc2>=0,1,0)
z0 <- alphaG + betaG*Xc2 - 0.2*statut
p0 <- 1/(1+exp(-z0))
z1 <- alphaD + betaD*Xc2 - 0.2*statut
p1 <- 1/(1+exp(-z1))
y1 <- rbinom(nTaille, 1, p1)
y0 <- rbinom(nTaille, 1, p0)
y <- ifelse(T_i==1,y1,y0)
df2 <- data.frame(Xc2, T_i, y, statut)
tt <- df2 %>%
filter(Xc2 %in% (-183:183))
fit <- glm(y~Xc2*T_i, family = "binomial", data=tt)
}

################################################
# 3. Simulation bayesienne par stan pour la
```

```
# loi a priori informative
################################################
## bibliotheques a utiliser
library(rstan)
## facilite le roulement des simulations
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
################################################
# 3.1 Code Stan
# le code qui suit doit etre inscrit dans un
# fichier separe: nomfichier.stan
################################################
# arguments
# N : nombre d'observations
# inter : variable d'interaction
################################################
data {
int<lower=0> N;
int<lower=0,upper=1> y[N];
vector [N] T_i;
vector [N] Xc2;
}
transformed data {
vector[N] inter;
inter   = T_i .* Xc2;
}
parameters {
real beta0;
real betaTi;
```

```
real betaXc2;
real betaInter;
}
model {
beta0 ~ normal(-3.178,0.5);
betaXc2 ~ normal(0,10);
betaInter ~ normal(0,14.14214);
betaTi ~ normal(-0.5913,0.7071068);
y ~ bernoulli_logit(beta0 + betaTi * T_i
+ betaXc2 * Xc2 + betaInter * inter);
}
################################################
# 3.2 Fonction de simulation bayesienne
################################################

flit_once <- function(x) {
gc()
gc()
df <- SimulationData(500,-3.178,0.00162,-2.5867,0.00162)
data.list <- list(N = nrow(df), y = df$y, T_i = df$T_i,
Xc2 = df$Xc2)
fitStan <- sampling(m, data = data.list, chains = 1,
 iter = 10000,seed=277360, warmup = 1000)
bh_summary <- summary(fitStan)$summary %>%
as.data.frame() %>%
mutate(variable = rownames(.)) %>%
dplyr::select(variable, everything()) %>%
as_data_frame() %>%
filter(variable=="betaTi") %>%
```

```
dplyr::select(c("mean","sd"))
}

# appelle la fonction pour rouler nos simulation
# extrait les resultats directement dans une base de donnees
mystan <- as.data.frame(sapply(1:1000,
FUN = function(r) {flit_once(r)}))

##############################################
# 3.3 Prediction a posteriori
##############################################
## arguments:
# Xc2: variable d'affectation
# T_i : variable indicatrice du traitement
# y : reponse binaire
# p0 : probabilite de succes quand T_i=0
# p1 : probabilite de succes quand T_i=1
# alphaD = -3.178
# betaD = 0.00162
# alphaG = -2.5867
# betaG = 0.00162
##############################################

ext_fit <- fitStan %>%
rstan::extract()
df$inter <- df$Xc2*df$T_i
# Extract posteriod distributions
alphaG_post <- ext_fit$beta0
betaG_post <- ext_fit$betaXc2
```

```
tau_post <-ext_fit$betaTi
inter_post <- ext_fit$betaInter

# Fonction de simulation en utilisant les memes variables
# d'affectation

gen_quant_r <- function(x,ti,interA) {
lin_comb <- sample(alphaG_post, size = length(x), replace = T) +
x*sample(betaG_post,size = length(x), replace = T) +
ti*sample(tau_post, size = length(ti), replace = T) +
interA*sample(inter_post, size = length(interA), replace = T)
prob <- 1/(1 + exp(-lin_comb))
out <- rbinom(length(x), 1, prob)
return(out)
}

# extrait les valeurs predites pour loi a priori informative
y_predInf <- as.data.frame(gen_quant_r(df$Xc2,df$T_i,df$inter))
colnames(y_predInf) <- "yPredI"
```

RÉFÉRENCES

Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. Récupéré le 2019-09-24 de https://arxiv.org/pdf/1701.02434.pdf

Black, D., Galdo, J. et Smith, J. (2003). Evaluating the regression discontinuity using experimental data.

Bor, J., Moscoe, E., Mutevedzi, P., Newell, M.-L. et Bärnighausen, T. (2014). Regression discontinuity designs in epidemiology causal inference without randomized trials. *Epidemiology (Cambridge, Mass.)*, *25*. http://dx.doi.org/10.1097/EDE.0000000000000138

Cerulli, G. (2015). Local average treatment effect and regression-discontinuity-design. In *Econometric Evaluation of Socio-Economic Programs : Theory and Applications* 229–305. Springer.

Ciolino, J., Martin, R., Zhao, W., Md, E., Hill, M. et Y Palesch, Y. (2013). Covariate imbalance and adjustment for logistic regression analysis of clinical trial data. *Journal of biopharmaceutical statistics*, *23*, 1383–402. http://dx.doi.org/10.1080/10543406.2013.834912

Defossez, A. (2019). Markov chain monte carlo (mcmc) and hybrid monte carlo (hmc). Récupéré le 2019-09-22 de https://ai.honu.io/presentations/mcmc_hmc.pdf

Denwood, M. J. (2016). runjags : An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, *71*(9), 1–25. http://dx.doi.org/10.18637/jss.v071.i09

Depaoli, S., Clifton, J. et Cobb, P. (2016). Just another gibbs sampler (jags) : Flexible software for mcmc implementation. *Journal of Educational and Behavioral Statistics*, *41*. http://dx.doi.org/10.3102/1076998616664876

Diaz-Quijano, F. (2012). A simple method for estimating relative risk using logistic regression. *BMC medical research methodology*, *12*, 14. http://dx.doi.org/10.1186/1471-2288-12-14

E Elder, T. (2010). The importance of relative standards in adhd diagnoses : Evidence based on exact birth dates. *Journal of health economics*, *29*, 641–56. http://dx.doi.org/10.1016/j.jhealeco.2010.06.003

Gabry, J. (2019). Graphical posterior predictive checks using the bayesplot package. https://mc-stan.org/bayesplot/articles/graphical-ppcs.html.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. et Rubin, D. B. (2014). Introduction to multiparameter models. In *Bayesian Data Analysis* chapitre 3, 63–82. CRC Press : Taylor Francis Group, (3 éd.).

Gelman, A., Carlin, J. B., Stern, H. S. et Rubin, D. B. (2004). *Bayesian Data Analysis* (2nd ed. éd.). Chapman and Hall/CRC.

Geneletti, S., O'Keeffe, A. G., Sharples, L. D., Richardson, S. et Baio, G. (2015). Bayesian regression discontinuity designs : Incorporating clinical knowledge in the causal analysis of primary care data. *Statistics in Medicine*, *34*(15), 2334–2352. http://dx.doi.org/10.1002/sim.6486

Geneletti, S., Ricciardi, F., O'Keeffe, A. et Baio, G. (2016). Bayesian modelling for binary outcomes in the regression discontinuity design.

Girolami, M. et Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society Series B*, *73*, 123–214.

Grenon-Godbout, N. (2015). Méthode mcmc : amélioration d'un algorithme d'adaptation régionale et applications à la climatologie. Récupéré le 2019-08-27 de https://dms.umontreal.ca/~bedard/GrenonG_rapport_final.pdf

Guo, J. et Geng, Z. (1995). Collapsibility of logistic regression coefficients. *Journal of the Royal Statistical Society : Series B (Methodological)*, *57*, 263–267. http://dx.doi.org/10.1111/j.2517-6161.1995.tb02029.x

Halldner, L., Tillander, A., Lundholm, C., Boman, M., Långström, N., Larsson, H. et Lichtenstein, P. (2014). Relative immaturity and adhd : Findings from nationwide registers, parent- and self-reports. *Journal of Child Psychology and Psychiatry*, *55*. http://dx.doi.org/10.1111/jcpp.12229

Hare, C. (2014). Applied bayesian modeling : A brief r2winbugs tutorial. Récupéré le 2020-01-27 de https://spia.uga.edu/faculty_pages/rbakker/bayes/Day3/R2winbugs.pdf

Hernán, M. et Robins, J. (2019). *Causal Inference.* Boca Raton : Chapman and Hall/CRC.

Imbens, G. W. et Lemieux, T. (2008). Regression discontinuity designs : A guide to practice. *Journal of Econometrics*, *142*(2), 615–635. http://dx.doi.org/10.1016/j.jeconom.2007.05.001

J D Barros, A. et Hirakata, V. (2003). Alternatives for logistic regression in cross-sectional studies : an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology*, *3*. http://dx.doi.org/10.1186/1471-2288-3-21

Jarvis, S. (2015). Introduction to winbugs with r. http://bes-qsig.github.io/fge/docs/IntroWinBUGSwithR/.

Joseph, L. (2019). Confounding and collinearity in multivariate logistic regression. Récupéré le 2019-08-28 de http://www.medicine.mcgill.ca/epidemiology/Joseph/courses/EPIB-621/logconfound.pdf

Krabbe, E., Thoutenhoofd, E., Conradi, M., Pijl, S. et Batstra, L. (2014). Birth month as predictor of adhd medication use in dutch school classes. *European Journal of Special Needs Education*, *29*(4), 571–578. http://dx.doi.org/10.1080/08856257.2014.943564. Récupéré de https://doi.org/10.1080/08856257.2014.943564

Kruschke, J. K. (2010). *Doing Bayesian Data Analysis : A Tutorial with R and BUGS* (1st éd.). USA : Academic Press, Inc.

L Morrow, R., Garland, E., Wright, J., Maclure, M., Taylor, S. et Dormuth, C. (2012). Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, *184*, 755–62. http://dx.doi.org/10.1503/cmaj.111619

Lee, J. K. et Chia, K. S. (1993). Estimation of prevalence rate ratios for cross sectional data : an example in occupational epidemiology. *British journal of industrial medicine*, *50*(9), 861–8622. http://dx.doi.org/10.1136/oem.50.9.861

Logan, M. (2018). Tutorial 7.5b - analysis of covariance (bayesian). http://www.flutterbys.com.au/stats/tut/tut7.5b.html#h2_16.

Lunn, D., Thomas, A., Best, N. et Spiegelhalter, D. (2000a). Winbugs - a bayesian modeling framework : Concepts, structure and extensibility. *Statistics and Computing*, *10*, 325–337.

http://dx.doi.org/10.1023/A:1008929526011

Lunn, D. J., Thomas, A., Best, N. et Spiegelhalter, D. (2000b). Winbugs a bayesian modelling framework : Concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325–337. http://dx.doi.org/http://dx.doi.org/10.1023/A:1008929526011

Matzke, D., Boehm, U. et Vandekerckhove, J. (2017). Bayesian inference for psychology, part iii : Parameter estimation in nonstandard models. *Psychonomic Bulletin Review*, *25*. http://dx.doi.org/10.3758/s13423-017-1394-5

Mazet, V. (2003). Introduction aux méthodes de monte-carlo par chaînes de markov. Récupéré le 2019-08-27 de http://miv.u-strasbg.fr/mazet/publis/mazet_mtde03.pdf

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. et Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092. http://dx.doi.org/10.1063/1.1699114

Morgan, S. L. et Winship, C. (2007). The counterfactual model. In *Counterfactuals and Causal Inference : Methods and Principles for Social Research* p. 31–58. Cambridge University Press.

N Evans, W., S Morrill, M. et Parente, S. (2010). Measuring inappropriate medical diagnosis and treatment survey data : The case of adhd among school-aged children. *Journal of health economics*, *29*, 657–73. http://dx.doi.org/10.1016/j.jhealeco.2010.07.005

O'Keeffe, A. G. et Baio, G. (2016). Approaches to the estimation of the local average treatment effect in a regression discontinuity design. *Scandinavian Journal of Statistics*, *43*, 978–995. http://dx.doi.org/doi:10.1111/sjos.12224

Oldenburg, C. E., Moscoe, E. et Bärnighausen, T. (2016). Regression discontinuity for causal effect estimation in epidemiology. *Current Epidemiology Reports*, *3*, 233–241. http://dx.doi.org/10.1007/s40471-016-0080-x

Ovando, D. (2018). Fitting bayesian modelling using stan and r. https://www.weirdfishes.blog/blog/fitting-bayesian-models-with-stan-and-r/.

Plummer, M. (2003). Jags : A program for analysis of bayesian graphical

models using gibbs sampling.

Plummer, M. (2019). *rjags : Bayesian Graphical Models using MCMC.* R package version 4-10

Porter, J. (2003). Estimation in the regression discontinuity model.

Potvin, J.-Y. (2019). Modèles stochastiques : Chaînes de markov discrètes.

R Core Team (2017). *R : A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria

Robinson, L. D. et Jewell, N. P. (1991). Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review, 59*, 227–240.

Shedden, K. (2019). Introduction to statistical computing : Confidence intervals. Récupéré le 2019-09-22 de `http://dept.stat.lsa.umich.edu/~kshedden/Courses/Stat406/Notes/confidence_intervals.pdf`

Stan Development Team (2018). The Stan Core Library. Version 2.18.0. Récupéré de `http://mc-stan.org/`

Stan Development Team (2019). RStan : the R interface to Stan. R package version 2.19.2. Récupéré de `http://mc-stan.org/`

Sturtz, S., Ligges, U. et Gelman, A. (2005). R2winbugs : A package for running winbugs from r. *Journal of Statistical Software, 12*(3), 1–16. Récupéré de `http://www.jstatsoft.org`

Su, Y.-S. et Yajima, M. (2015). *R2jags : Using R to Run 'JAGS'.* R package version 0.5-7

Surhone, L. M., Tennoe, M. T. et Henssonow, S. F. (2010). *OpenBUGS.* Beau Bassin, MUS : Betascript Publishing.

Thistlethwaite, D. L. et Campbell, D. T. (1960). Regression-discontinuity analysis : An alternative to the ex post facto experiment.

Turgeon, M. (2017). *Portrait de l'usage des médicaments spécifiques au trouble du déficit de l'attention avec ou sans hyperactivité (TDAH) chez les Québécois de 25 ans et moins. Revue systématique sur les recommandations de bonne pratique quant à l'usage optimal des médicaments spécifiques au traitement du trouble du déficit de l'attention avec ou sans hyperactivité (TDAH) et annexes complémentaires.* Publications gouvernementales du Québec en ligne : monographies électroniques. Récupéré le 2019-08-27 de `http://collections.banq.qc.ca/ark:/52327/3109470`

Williamson, T., Eliasziw, M. et Hilton Fick, G. (2013). Log-binomial models : Exploring failed convergence. *Emerging themes in epidemiology*, *10*, 14. http://dx.doi.org/10.1186/1742-7622-10-14

Xing, G. et Xing, C. (2010). Adjusting for covariates in logistic regression models. *Genetic epidemiology*, *34*, 769–71 ; author reply 772. http://dx.doi.org/10.1002/gepi.20526

Zou, G. (2004). A modified poisson regression approach to prospective binary data. *American journal of epidemiology*, *159*, 702–6. http://dx.doi.org/10.1093/aje/kwh090