Rapport de stage

Déserts médicaux

Auteur : Ghiles Kemiche

Nom de l'organsme d'accueil : CREST-ENSAE

Maître De Stage : Pr. Guillaume Lecué

Licence double diplôme Mathématiques - Informatique

2022 - 2023

Remerciements

Je souhaite exprimer mes sincères remerciements à mon maître de stage, Monsieur Guillaume Lecué, professeur à l'ENSAE/ESSEC, pour m'avoir accordé sa confiance et m'avoir accueilli en tant que stagiaire. Sa disponibilité et son professionnalisme irréprochable ont été d'une grande valeur tout au long de mon stage.

Je suis reconnaissant de sa générosité en partageant son expertise et ses précieux conseils, qui ont grandement contribué à mon apprentissage. Son accompagnement attentif m'a permis de mieux comprendre le domaine et d'approfondir mes connaissances.

De plus, j'apprécie sincèrement l'environnement de travail sain et chaleureux qu'il a su créer. Sa bienveillance et son ouverture d'esprit ont favorisé une atmosphère propice à l'épanouissement personnel et professionnel.

Je tiens également à souligner son rôle dans l'orientation de mes études. Ses conseils éclairés ont été d'une grande aide pour définir mes objectifs académiques et professionnels, me permettant ainsi d'aborder mon avenir avec plus de clarté et de confiance.

En résumé, je suis extrêmement reconnaissant envers Monsieur Guillaume Lecué pour son accueil, son encadrement et son soutien tout au long de mon stage. Sa bienveillance, son expertise et ses conseils précieux ont été des atouts indispensables dans ma formation et mon développement professionnel.

(G)

Table des matières

Table des figures

1

Mission et étapes

1.1 Contexte

Au sein de ce projet, notre objectif est de développer des indicateurs et des estimateurs qui permettent de mesurer les notions "d'offre médicale", de "demande de soins de santé" et de densité de l'offre médicale. L'accent est particulièrement mis sur l'utilisation de ces outils pour identifier les zones où l'accès aux soins médicaux est limité, communément appelées "déserts médicaux".

1.2 Modèle Proposé

L'approche naïve et naturelle pour mesurer la densité médicale consiste à diviser le nombre de médecins d'une zone géographique par le nombre de sa population. Cependant, cette approche ne prend pas en compte le fait qu'une personne peut se rendre chez un médecin situé dans une autre zone géographique, la concurrence entre prestataires, etc. Afin de remédier à cette limitation, nous proposons trois approches basés sur les modèles xSFCA (x Steps Floating Catchment Area), qui prennent en considération cette problématique.

Nos modèles se basent sur les données suivantes :

- P_i : représente la demande de la zone i, c'est-à-dire les besoins en soins de santé de la population résidant dans cette zone.

- W_{ij} : correspond au coefficient de perméabilité de la demande de la zone i pour se rendre dans la zone j. Ce coefficient peut être l'inverse de la distance entre la zone i et la zone j : $1/d_{ij}$ ou une fonction dépendant de cette distance.

- S_j : est l'offre de soins de santé de la zone j, représentant la mesure des soins de santé fournis par l'ensemble des prestataires de soins de santé de cette zone.

Le but de ces données est de déterminer $\mathbb{P}_{ij} = \mathbb{P}[i \sim j]$, qui est la probabilité que les patients de la zone i consultent des prestataires de soins de santé dans la zone j. La définition de cette quantité varie d'une approche à une autre. À partir de là, nous construisons les mesures suivantes :

- l'offre totale de la zone j par patient potentiel : $R_j = \frac{S_j}{\sum_i P_i \mathbb{P}_{ij}}$

- l'offre médicale totale desservie dans la zone i : $FCA_i = \sum_j \mathbb{P}_{ij} R_j$

Ainsi, nous pouvons définir un désert médical en utilisant l'offre médicale totale desservie. En effet, si FCA_i pour la zone i se situe parmi les 10% les plus faibles des FCA_i (ie quantile d'ordre $1/10$), alors nous pouvons considérer la zone i comme un désert médical.

1.2.1 Modèle 2SFCA (1)

Cette méthode ne tient compte que de la distance de l'offre. La méthode 2SFCA utilise les définitions suivantes pour W_{ij} et $\mathbb{P}_{ij}$:

$$W_{ij} = \frac{1}{d_{ij}} \quad \text{et} \quad \mathbb{P}_{ij} \propto W_{ij} \;\; \text{i.e.} \;\; \mathbb{P}_{ij} = \frac{W_{ij}}{\sum_k W_{ik}}$$

où d_{ij} est la distance entre i et j.

1.2.2 Modèle 3SFCA (1)

Cette méthode tient compte de la distance de l'offre et l'offre de soin. La méthode 3SFCA utilise les définitions suivantes pour W_{ij} et $\mathbb{P}_{ij}$:

$$W_{ij} = \frac{1}{d_{ij}} \quad \text{et} \quad \mathbb{P}_{ij} \propto \mathbb{P}_{ij} \propto W_{ij} S_j \;\; \text{i.e.} \;\; \mathbb{P}_{ij} = \frac{W_{ij} S_j}{\sum_k W_{ik} S_k}$$

où d_{ij} est la distance entre i et j.

1.2.3 Modèle du point fixe ou ∞SFCA (2)

Après avoir introduit les méthodes 2SFCA et 3SFCA, il est naturel d'itérer le processus. L'idée principale derrière la méthode de point fixe FCA, également appelée $\infty SFCA$, est que la probabilité $\mathbb{P}_{ij}$, dépend du à la fois du coefficient de perméabilité W_{ij} et de l'offre totale de la zone j par patient potentiel R_j.

Dans la méthode de point fixe SFCA, on utilise les définitions suivantes pour W_{ij} et $\mathbb{P}_{ij}$:

$$W_{ij} = \frac{1}{d_{ij}} \quad \text{et} \quad \mathbb{P}_{ij} \propto \mathbb{P}_{ij} \propto W_{ij}R_j \;\; \text{i.e.} \;\; \mathbb{P}_{ij} = \frac{W_{ij}R_j}{\sum_k W_{ik}R_k}$$

Cependant, le problème avec cette méthode est que la définition de R_j dépend elle-même des valeurs de $(\mathbb{P}_{ij})_i$. Ainsi, on a les équations suivantes pour trouver ces deux quantités :

$$R_j = \frac{S_j}{\sum_i P_i \mathbb{P}_{ij}} \quad \text{et} \quad \mathbb{P}_{ij} = \frac{R_j W_{ij}}{\sum_k R_k W_{ik}}$$

Ce qui est équivalent à trouver le point fixe de la fonction suivante :

$$F : \begin{cases} \mathbb{R}^J \times \mathbb{R}^{N\times J} & \rightarrow \mathbb{R}^J \times \mathbb{R}^{N\times J} \\ \left(R = (R_j)_j \,, \mathbb{P} = (\mathbb{P}_{ij})_{ij}\right) & \mapsto (F_1(\mathbb{P}), F_2(R)) \end{cases}$$

où $F_1(\mathbb{P}) = \left(\frac{S_j}{\sum_i P_i \mathbb{P}_{ij}}\right)_j$ et $F_2(R) = \left(\frac{R_j W_{ij}}{\sum_k R_k W_{ik}}\right)_{ij}$

Pour trouver le point fixe de cette fonction nous allons utiliser l'algorithme du point fixe qui consiste à calculer pour chaque iteration :

$$\mathbb{P}^{(k+1)} = F_1\left(R^{(k)}\right) \text{ and } R^{(k+1)} = F_2\left(\mathbb{P}^{(k+1)}\right)$$

Pour Etudier la convergence de l'algorithme du point fixe on decide d'etudier la convergence vers 0 des suites $\|R^{(k+1)} - R^{(k)}\|$ et $\|\mathbb{P}^{(k+1)} - \mathbb{P}^{(k)}\|$

1.3 Distancier

Pour obtenir les distances $d_{i,j}$ entre chaque paire de zones géographiques (i, j), nous utiliserons le distancier Metric de l'INSEE. Il s'agit d'un logiciel de requête auquel il est nécessaire de faire une demande à la chaire Quetelet pour y accéder. Cependant, le distancier présente quelques inconvénients. Tout d'abord, il est limité à la France métropolitaine, ce qui signifie qu'il ne couvre pas les autres territoires. De plus, les requêtes peuvent prendre un temps considérable. Par exemple, requêter les distances entre chaque commune de France prendrait environ 157 heures. Il existe également d'autres distanciers tels que Odamatrix et celui de l'IRDES, mais nous n'avons pas accès à ceux-ci.

1.4 Offre et demande de soin : $(S_j)_j$ et $(P_i)_i$

1.4.1 Première approche

Une approche naïve consiste à considérer que l'offre de soins pour une zone j (S_j) est le nombre de médecins, et la demande de soins (P_i) est le nombre de personnes résidant dans cette zone. Pour obtenir ces données, nous avons décidé d'utiliser les bases de données de l'INSEE : ici pour les P_i et ici pour les S_j.

Ces bases de données utilisent un code géographique unique attribué à chaque commune. Cependant, elles présentent certaines incompatibilités entre elles. Par exemple, Paris est considéré comme une commune dans l'une des bases, tandis que dans l'autre base, ses arrondissements sont représentés comme des communes distinctes.

1.4.2 Effets fixes

La première approche ne quantifie pas de manière précise l'offre réelle et la demande réelle de soins, car tous les patients n'ont pas le même effet. Par exemple, une personne âgée aura une demande beaucoup plus importante qu'un jeune, de même certains médecins peuvent avoir un nombre élevé de consultations par rapport à d'autres (pour des raisons économiques ou en raison d'une pénurie de médecins dans leur région, par exemple). Il est donc important de prendre en compte les effets fixes de chaque patient et de chaque médecin afin d'estimer de manière plus précise l'offre et la demande de soins. Cette approche consiste à trouver des effets fixes pour chaque patient et chaque médecin, en formant un graphe biparti où les médecins d'un côté et les patients de l'autre sont reliés par des liens, où la présence d'un lien signifie qu'une consultation a eu lieu. Ce modèle est appelé AKM (Abowd, Kramarz, et Margolis).

1.4.2.1 Modèle logit

L'idée est d'estimer le modèle AKM dans le cas d'une variable binaire Y. Le modèle peut être écrit de la manière suivante :

$$\ln \frac{\mathbb{P}\left[Y_{ij} = 1 \mid X_{ij}\right]}{\mathbb{P}\left[Y_{ij} = 0 \mid X_{ij}\right]} = X_{ij}\beta + \alpha_i + \psi_j + \epsilon_{ij}$$

où :

- Y_{ij} : la liaison entre le patient i et le médecin j, qui vaut 1 si la liaison existe, sinon 0.
- α_i : l'effet fixe du patient i.
- ψ_j : l'effet fixe du médecin j.
- X_{ij} : la matrice des caractéristiques observables de la paire patient-médecin (i, j).
- β : le vecteur de coefficients de régression à estimer.
- ϵ_{ij} : l'erreur résiduelle.

On peut remarquer que $\mathbb{P}\left[Y_{ij} = 1 \mid X_{ij}\right] = \sigma(X_{ij}\beta + \alpha_i + \psi_j + \epsilon_{ij})$ où σ est la fonction sigmoid

Le but est d'utiliser une base de données $(Y_{ij}, X_{ij})_{ij}$ pour estimer les effets fixes de chaque patient et de chaque médecin, ainsi que le vecteur de paramètres β. Une fois ces estimations obtenues, on peut se placer dans une zone géographique donnée. Ainsi, la demande totale de soins dans cette zone est égale à la somme des effets fixes des patients qui y résident, et de même, l'offre totale de soins dans cette zone est égale à la somme des effets fixes des médecins qui y résident. Cela permet de quantifier de manière précise les $(S_j)_j$ et $(P_i)_i$.

Références

[1] L.-G. Véronique et Mangerey. Inégalités spatiales d'accessibilité aux médecins spécialistes. Technical report, Irdes et ORS Ile-de-France, 2022. iii, 2

[2] G. L. e. Indicators of medical deserts. Technical report, CREST, ENSAE, IPParis, 2023. iii, 3