



HAL
open science

Modélisation des données d'attractivité hospitalière par les modèles d'utilité

Issa Saley

► **To cite this version:**

Issa Saley. Modélisation des données d'attractivité hospitalière par les modèles d'utilité. Autres [stat.ML]. Université Montpellier; Université Abdou Moumouni, 2017. Français. NNT : 2017MONT044 . tel-01926375

HAL Id: tel-01926375

<https://theses.hal.science/tel-01926375>

Submitted on 19 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de
Docteur

Délivré par les **Universités de Montpellier**
et **Abdou Moumouni de Niamey**

Préparée au sein des écoles doctorales **I2S** et **EDSET**
Et des unités de recherche **IMAG** et **FAST**

Spécialité: **Biostatistique**

Présentée par **Issa SALEY**

Modélisation des données d'attractivité hospitalière par les modèles d'utilité

Soutenue le 29 Novembre 2017 devant le jury composé de

Ahmadou ALIOUM	Pr	Université de Bordeaux	Président/Rapporteur
Diakarya BARRO	Pr	Université de Ouaga 2	Rapporteur
Nicolas MOLINARI	Pr	Université de Montpellier	Directeur de thèse
Bisso SALEY	Pr	Université de Niamey	Co-Directeur
Mathieu RIBATET	M. Conf.	Université de Montpellier	Invité
Bénédicte FONTEZ	M. Conf.	Université de Montpellier	Invité



Remerciements et Dédicaces

Je commence tout d'abord par remercier le Professeur Nicolas Molinari, sans qui, probablement cette thèse n'aurait pas eu lieu. Il a accepté volontiers de m'accompagner dans le processus de la recherche du financement de cette thèse au niveau de l'ambassade de France au Niger et de diriger mon travail malgré qu'il a y a des années que j'ai quitté les bancs. Il m'a fait confiance et m'a donné une liberté totale. C'est le lieu de lui témoigner toute ma reconnaissance. Je remercie également le Professeur Issoufou Kouada qui malgré une santé chancelante a accepté de co-diriger cette thèse ; santé chancelante qui l'a contraint à prendre une retraite anticipée avant la fin de mon travail. Je prie Dieu pour qu'il recouvre sa santé. Mes remerciements et ma reconnaissance vont également à l'endroit du Professeur Saley Bisso qui a pris le relais de Kouada sans hésiter et m'a beaucoup facilité le coté administratif à l'Université de Niamey. Merci à Mathieu Ribatet, en qui j'ai trouvé un co-encadrant rigoureux et talentueux, même si les conditions de parité des membres du jury entre les Universités de Montpellier et Niamey ont fait qu'il apparait comme membre invité du jury. Je remercie les Professeurs Ahmadou Alioum et Diakarya Barro qui ont évalué ce travail. J'espère avoir tenu compte au mieux dans mon manuscrit de leurs remarques et suggestions pertinentes. Je remercie aussi Bénédicte Fontez, membre invité du jury qui a été membre de tous mes comités de suivi de thèse avec des remarques très instructives. Mon premier clin d'œil va à Yann Cabon, doctorant de Nicolas pour des échanges enrichissants dans notre bureau au DIM. C'est le lieu de remercier également toute l'équipe de DIM et principalement la secrétaire Sophie Maurin pour mes incessantes sollicitations. Aussi je profite pour remercier du coté de la représentation de l'ASECNA au Niger dont je suis agent, Elhadj Hassane Hami, Chef d'antan de mon unité, Ali Tsayabou, le Responsable d'opérations à l'époque et Lamine Garba, le Représentant de l'agence au Niger pour avoir été compréhensifs et flexibles administrativement à mon endroit ; clin d'œil à tous les membres de l'équipe unité Télécoms.

A défaut de les citer nommément et pour ne pas faire de jaloux, je remercie tous

mes amis (et Dieu sait qu'ils sont nombreux) partout ailleurs et leur témoigne toutes mes gratitudes pour leur soutien de tous genres.

Dédicace à ma femme Rabi dite Nadine à qui j'ai très souvent manqué durant cette thèse et à ma fille Nesrine-Rabi née durant cette thèse et que je n'ai pu voir que six mois après sa naissance. Dédicace à mon oncle Amadou Garka chez qui j'ai fait une partie de mes études mais qui est décédé avant la fin de cette thèse. Enfin, spéciale dédicace à mes frères et sœur, à mon père décédé juste quelques deux mois après que j'ai commencé à aller à l'école et de qui je n'ai qu'une vague image et à ma mère Rabi qui a tout fait pour moi : j'aurai souhaité qu'elle soit là lors de la soutenance de son fils-cadet mais elle aussi est décédée quelques mois avant que je débute cette thèse.

Table des matières

1	Introduction	1
2	Outils	5
2.1	Les modèles de choix discret	5
2.1.1	Les modèles à Règle de Décision Aléatoire	6
2.1.2	Les modèles d'Utilité Aléatoire (RUMs)	10
2.2	Géostatistique : le processus gaussien	13
2.2.1	Champ gaussien	14
2.2.2	Exemples de fonctions de covariance isotropes	14
2.3	Éléments de la Statistique Bayésienne	16
2.3.1	Modélisation bayésienne	16
2.3.2	Inférence bayésienne	17
2.3.3	Choix de modèles	18
2.3.4	Remarques	19
2.4	Méthodes de Monte Carlo par Chaînes de Markov	19
2.4.1	Définition	19
2.4.2	Échantillonnage de Gibbs	21
2.4.3	Algorithme Hastings-Metropolis	21
2.4.4	Remarques	22
3	Effet de la distance sur le nombre d'admission de patients au niveau d'un hôpital	23
3.1	Les modèles avec des extra-zéros	24
3.1.1	Le ZIP et le Hurdle	24
3.1.2	Le modèle ZINB	25
3.2	Application de ces modèles aux données d'admission de patients	26
3.2.1	Cadre et précisions	26
3.2.2	Modélisation des données d'admission de patients	26

3.2.3	Amélioration de la modélisation des données par prise en compte de la dépendance spatiale	27
3.3	Article	29
4	Effet de la distance sur la probabilité de choix d'un hôpital	45
4.1	Application de ces modèles aux données d'admission de patients	45
4.1.1	cadre et précisions	45
4.1.2	Modélisation des données d'admission de patients	46
4.1.3	Prise en compte de la corrélation spatiale	46
4.2	Article	47
4.3	Annexe	58
5	Envergure territoriale d'un hôpital	61
5.1	Application d'un RUMs : le Logit	61
5.1.1	modèle	62
5.1.2	Estimation des paramètres	63
5.1.3	résultats	64
5.1.4	Limites des modèles RUMs au vu de notre but	67
5.2	Une autre approche pour calculer l'envergure	68
5.2.1	L'approche proposée	68
5.2.2	Application aux données d'admission de patients	69
5.3	Article	70
6	Conclusion et Perspectives	87
A	Mémento sur les chaînes de Markov	89
A.1	Généralités	89
A.2	Propriétés	90
A.2.1	Homogénéité	90
A.2.2	Irréductibilité	90
A.2.3	Apériodicité	91
A.2.4	Réurrence	91
A.3	loi stationnaire	91
A.4	Ergodicité et convergence	91
B	Quelques fonctions et distributions	92
B.1	Fonction gamma (Γ) -Fonction de Bessel modifiée	92
B.1.1	Fonction Gamma (Γ)	92
B.1.2	Fonction de Bessel modifiée	92
B.2	Quelques distributions	93
B.2.1	Loi gaussienne et loi log-normale	93
B.2.2	Loi Inverse-gamma	94
B.2.3	Loi de Poisson- Loi Binomiale négative	94

Table des figures

1.1	Localisation de la région de Languedoc-Roussillon avec ses cinq départements sur la carte de la France	3
4.1	Le graphe acyclique direct du modèle	58
5.1	Chaine MCMC et densité conditionnelle du paramètre effet distance pour le CH de Perpignan	65
5.2	Chaine MCMC et densité conditionnelle du paramètre effet distance pour la Polyclinique Saint-Roch	66

Liste des tableaux

5.1	Nombre de séjours enregistrés dus à l'asthme par chacun des deux hôpitaux en 2009	62
5.2	Table des valeurs moyennes des β_j et leur intervalle de crédibilité à 95%	67

CHAPITRE 1

Introduction

Dans le cadre de la réforme du système de santé français, une loi dite Loi de Tarification à l'Activité (T2A) a été mise en place en 2005 par les autorités. Elle base la rémunération des établissements de santé (publics et privés) sur leurs activités médicales, instaurant de fait une concurrence entre les établissements de santé publics et aussi entre publics et privés s'ajoutant à celle déjà existant entre établissements privés. Les établissements adoptent des mesures et politiques pour qu'ils soient plus attractifs pour les patients afin d'avoir plus de revenus. Attractif s'entend par la qualité des soins, le coût, l'accueil, les commodités. . .

Notre travail de thèse a consisté à modéliser des données d'admission de patients afin de prévoir le flux des patients et/ou de comparer l'attractivité des hôpitaux et donc d'en établir une classification. Pour ce faire, la distance du lieu de résidence du patient par rapport à l'hôpital a été étudiée comme modificateur de cette attractivité. En effet, même si la sensibilité à la distance peut dépendre de l'âge du patient et du type de maladie [Porell and Adams, 1995], des études [Lako and Rosenau, 2009] ont montré que la distance par rapport à l'hôpital influe très négativement sur le choix du patient, surtout dans une région où on peut trouver les hôpitaux à des distances relativement courtes [Victoor et al., 2012]. Donc nous avons estimé que si un malade peut faire une longue distance pour se rendre dans un hôpital au lieu de ceux plus proches, c'est que cet hôpital répond beaucoup plus à ses souhaits, donc plus attractif : accès facile, qualité des soins, coût, commodités. . . . Alors toutes les modélisations que nous sommes amené à faire dans ce travail sont faites autour de la distance parcourue par le patient pour se rendre dans un hôpital donné.

L'intérêt de cette modélisation/classification pour un établissement de soins serait double :

- 1) il peut utiliser cette classification pour valoriser son offre de soins auprès des patients dans un environnement concurrentiel et pour rechercher des finance-

ments dans le cadre la Loi de la Tarification à l'Activité et

- 2) il peut optimiser ses dépenses et des conditions d'accueil des patients sachant le flux.

De plus, pour les autorités sanitaires, elle aidera à l'allocation des ressources pour un financement optimal des activités hospitalières et les aider dans l'analyse ou l'implémentation de certaines reformes pour l'offre des soins de santé.

Pour la modélisation de ces données, au tout début de la formulation du sujet, l'idée était d'utiliser seulement les modèles d'Utilité Aléatoire (Random Utility Models ou RUMs en abrégé). En effet, le choix d'un hôpital par un patient pour admission peut être vu comme le choix d'un produit parmi tant d'autres ; on peut supposer qu'un patient choisit un hôpital selon la satisfaction qu'il lui procure. Mais dès l'entame de ce travail, nous nous sommes rendu compte, qu'outre les modèles d'Utilité Aléatoire, bien d'autres modèles peuvent être utilisés pour modéliser les données d'attractivité hospitalière selon l'angle que l'on considère pour l'attractivité.

Par exemple, elle peut être vue sous l'angle de la couverture territoriale, ou en terme du nombre moyen de patients admis ou bien de la probabilité que l'établissement soit choisi parmi tant d'autres. C'est ainsi que nous avons utilisé deux autres modèles pour ces données en plus des modèles d'Utilité Aléatoire. Il appartient donc à l'analyste de choisir la modélisation qui semble le mieux s'inscrire dans son approche parmi celles proposées.

Pour l'illustration, des données d'admission de patients dans la région Languedoc-Roussillon ont été utilisées dont la principale source est le Programme de médicalisation des systèmes d'information plus connu sous l'acronyme du *PMSI*. C'est un dispositif administratif où sont enregistrées toutes les admissions des patients dans le système hospitalier français, public et privé et quel que soit le régime d'assurance. Dans ce dispositif, des informations sur le patient (telles que son âge, son sexe, la pathologie, son lieu de résidence...) et sur l'hôpital d'admission (type d'hôpital, localisation...) sont recueillies. La région de Languedoc-Roussillon était la première à implémenter ce programme en France. C'est une région connue pour ses services de soins de haut niveau. Géographiquement, située dans le Sud de la France, avec une superficie de $27\,376\text{ km}^2$, elle s'étend de $1,5$ au 5 degrés Longitude Est et de 42 à $44,5$ degrés Latitude Nord. Elle est composée de 5 départements : l'Aude, le Gard, l'Hérault, le Lozère et les Pyrénées-Orientales avec pour chef lieu Montpellier. La figure 1 donne la localisation géographique de la région.

Elle compte environ 3 millions d'habitants dont pour à peu près 2 millions d'actifs en 2013 selon l'Institut national de la statistique et des études économiques. Plus de 16 pour 100 de la population est retraitée et elle a un des plus hauts taux d'accroissement en France : 1,1 pour 100 contre 0,4 pour 100 au niveau national.

Le travail s'articule autour de trois chapitres, en plus du premier chapitre consacré aux outils statistiques et algorithmes que nous avons utilisés ; des chapitres qui peuvent être lus indépendamment. Bien qu'il n'y ait pas de lien mathématique formel entre les différents modèles proposés, nous avons essayé de garder un agencement

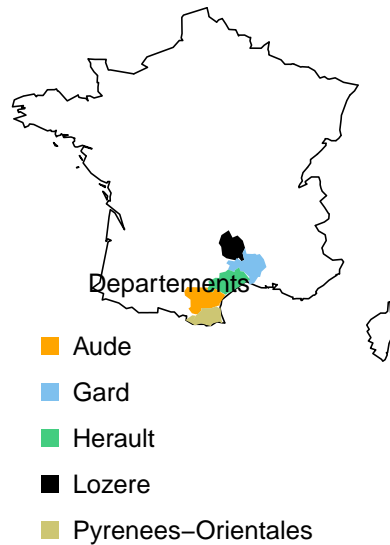


FIGURE 1.1 – Localisation de la région de Languedoc-Roussillon avec ses cinq départements sur la carte de la France

du travail de sorte que les chapitres soient dans l'ordre des modèles qui semblent plus proches dans l'approche : le second et le troisième chapitres sont des modèles de données de comptage avec corrélation spatiale même si le premier se centre sur le nombre d'admissions et le second sur la probabilité de choisir un établissement par localité, le troisième et le quatrième se rejoignent par le fait qu'ils sont centrés sur la probabilité de choix au niveau d'une localité, probabilité qui peut être même ramenée au niveau de l'individu même si ces modèles l'appréhendent de manière différente.

Dans le premier chapitre, la première section traite des modèles de choix discret et la seconde un peu du processus gaussien souvent utilisé dans la modélisation des données avec dépendance spatiale comme les nôtres. Du fait que toutes les modélisations dans cette thèse se sont faites dans un cadre bayésien paramétrique, la troisième section présente quelques éléments de statistique bayésienne. Comme en inférence bayésienne, les distributions a posteriori sont souvent de forme complexe et rendant pratiquement impossible tout calcul analytique, on a alors recours aux simulations. Les méthodes Monte Carlo par Chaînes de Markov sont généralement utilisées pour les simulations. C'est pourquoi dans la quatrième et dernière section de ce chapitre deux algorithmes que nous avons utilisés, à savoir l'échantillonneur de Gibbs et l'algorithme de Hastings-Metropolis, sont présentés.

Le chapitre 2 porte sur des modèles pour des données de comptage lorsque le nombre de zéros présents est différent de ceux attendus pour la loi utilisée. Dans un premier temps, une vue d'ensemble a été donnée sur trois de ces modèles. Ensuite,

une application sur des données d'admission due à l'asthme dans l'Hérault où le nombre moyen de patients d'une localité admis dans un hôpital est considéré dépendant de la distance de la localité par rapport à l'hôpital avec corrélation spatiale entre localités entrant dans l'étude, a été présentée.

L'utilisation du modèle binomial hiérarchique avec application sur des données de cancer de prostate pour deux hôpitaux de Nîmes (Gard) est donnée dans le chapitre 3. Avec ce modèle, la probabilité moyenne de choisir un hôpital dépend de la distance de la localité de résidence du patient par rapport à l'hôpital avec une possible dépendance spatiale des probabilités entre localités entrant dans l'étude.

Le quatrième chapitre porte sur les modèles d'utilité. Une application d'un modèle RUMs (Logit) est faite sur des données concernant les séjours hospitaliers dus à l'asthme dans les Pyrénées-Orientales en 2009. Compte tenu de certaines limites présentées par ces modèles au vu de notre objectif, une nouvelle approche est introduite. Avec cette approche, l'hypothèse selon laquelle un agent choisit entre produits en compétition celui qui lui procure le plus d'utilité dans le cas des RUMs est relâchée pour une hypothèse plus souple ; hypothèse selon laquelle l'agent peut choisir un produit dès lors que l'utilité que lui procure ce produit a atteint un certain seuil de satisfaction en considérant certains aspects. Une illustration de cette approche est faite sur trois hôpitaux de l'Hérault pour les séjours dus à l'asthme en 2009.

Une conclusion générale avec des perspectives du travail est donnée en fin des chapitres, suivie des annexes.

CHAPITRE 2

Outils

Ce chapitre a pour but de donner de manière succincte des outils de modélisation et inférence auxquels nous avons eu recours au cours de notre travail. Ainsi la section 2.1 présente les modèles de choix discret, avec un accent particulier sur les modèles d'Utilité Aléatoire. Comme nous l'avons dit en introduction, outre les modèles d'Utilité Aléatoire, d'autres modèles, prenant en compte une possible dépendance spatiale des données, peuvent être utilisés. De ce fait, la section 2.2 présente le processus gaussien qui est souvent utilisé directement ou indirectement en pareil cas. Et comme toutes nos inférences statistiques ont été conduites dans un cadre bayésien paramétrique, la section 2.3 présente de manière sommaire des éléments de la statistique bayésienne paramétrique. La section 2.4 porte sur la présentation de deux algorithmes des méthodes Monte Carlo par Chaînes de Markov les plus fréquemment utilisés dans l'inférence bayésienne : l'échantillonneur de Gibbs et l'algorithme de Hastings-Metropolis.

2.1 Les modèles de choix discret

Les modèles de choix discret sont des modèles qui décrivent le comportement de choix d'un individu devant choisir une alternative parmi un **ensemble fini** d'alternatives qui sont **mutuellement exclusives**. Par alternative, s'entend toute action, toute option ou tout objet sur lesquels devrait se porter une décision ou un choix. Contrairement à l'économie néoclassique qui suppose que l'individu dispose d'une relation de préférence pouvant être associée à une fonction d'utilité, lui permettant de classer les alternatives par ordre de préférence et de ce fait choisir celle qui lui procurerait la plus grande utilité, les modèles de choix discret utilisent plutôt une démarche probabiliste pour expliquer le choix de l'individu. Cette démarche tire son

fondement du fait des fluctuations que peuvent avoir les choix de l'individu en fonction du contexte. Le choix de l'individu est donc supposé soumis à un aléa ; ce qui fait que pour chaque alternative, sa probabilité d'être choisie par l'individu peut être calculée. Et selon le mécanisme aléatoire auquel serait soumis le choix, deux familles principales de modèles de choix discret sont distinguées :

1. Les modèles à Règle de Décision Aléatoire et
2. Les modèles d'Utilité Aléatoire,

familles que nous présentons dans les deux prochaines sous-sections.

2.1.1 Les modèles à Règle de Décision Aléatoire

Dans cette famille de modèles de choix discret, l'utilité que procurerait une alternative à un individu est supposée déterministe et le processus du choix de l'individu, aléatoire. L'individu ne choisit pas forcément l'alternative avec la plus grande utilité. Deux modèles font figure d'exemples dans cette famille ; les autres plus ou moins leurs variantes. Il s'agit du modèle de Luce [Luce, 1959] et celui de Tversky [Tversky, 1972]. Le premier est basé sur un axiome permettant de relier les probabilités de choix des alternatives entre elles et le second, sur une procédure. Avant l'exposition de ces modèles, il importe de donner les notations suivantes :

- E : ensemble fini des alternatives sur lesquelles s'opérerait le choix,
- \mathcal{S} : ensemble des parties de E constituées d'au moins 2 alternatives,
- $P_S(a)$: la probabilité que l'alternative a soit choisie parmi celles de la partie $S \in \mathcal{S}$.
Si S est exactement constituée de 2 alternatives, $P_S(a)$ est notée $P(a, b)$ où b est l'autre alternative.
- $P_E(S) = \sum_{a \in S} P_E(a)$

Le modèle de Luce

Pour construire son modèle, Luce commence d'abord par poser un axiome sur les liens qui doivent exister entre les probabilités de choix de différentes alternatives d'un ensemble fini E . Il suppose que pour toutes parties $S, T \in \mathcal{S}$ telles que $S \subset T$,

- (i) si pour n'importe quelle alternative $a \in E$ donnée, $P(a, b) \neq 0, 1$ pour toute alternative $b \in T$, alors

$$P_T(a) = P_T(S) \cdot P_S(a). \quad (2.1)$$

Cette première partie de l'axiome dit que la probabilité du choix d'une alternative dans T est indépendante des sous-ensembles de T qui la contiennent. Donc peu importe comment l'individu subdivise l'ensemble des alternatives pour opérer son choix, la probabilité de choisir une alternative reste la même dans cet ensemble. Elle a pour conséquence que le ratio des probabilités de choix de deux alternatives est indépendant des autres alternatives. Autrement dit, qu'on ajoute ou qu'on retire une alternative, les différents ratios restent

inchangés. En effet soient S et S' deux parties distinctes et toutes deux appartenant à \mathcal{S} . Il résulte de (i) que

$$\begin{aligned} P_T(a) &= P_T(S).P_S(a) = P_T(S').P_{S'}(a) \\ P_T(b) &= P_T(S).P_S(b) = P_T(S').P_{S'}(b) \end{aligned}$$

En divisant membre à membre, on obtient que

$$\frac{P_T(a)}{P_T(b)} = \frac{P_S(a)}{P_S(b)} = \frac{P_{S'}(a)}{P_{S'}(b)}.$$

Cette propriété est connue sous le nom de l'indépendance par rapport aux choix extérieurs ou en anglais, irrelevant from independent alternatives (IIA) ; un fait qui n'est pas tout à fait réaliste dans certaines situations et qui constitue un point faible du modèle. En témoigne un contre-exemple de Debreu [Debreu, 1960], bien connu : le paradoxe du bus bleu/bus rouge. Dans cet contre-exemple, l'idée est qu'on suppose qu'un individu pour se rendre à une destination soit indifférent au choix de bus ou sa propre voiture, c'est-à-dire que la probabilité d'utiliser sa voiture est la même que celle d'utiliser un bus, supposons bleu. Puis un second bus identique au premier, mais supposons de couleur rouge, est introduit. Pour une question de formalisme, désignons la voiture par a , le bus bleu par b et le bus rouge par r . Si l'individu est indifférent aux couleurs des bus (somme, logique), on est en droit de dire que la probabilité de choisir la voiture par rapport au bus bleu est égale à celle de choisir la voiture par rapport à celui rouge, c'est-à-dire que $P(a, b) = P(a, r)$. Aussi on a la probabilité de choisir le bus rouge est la même que celle de choisir le bus bleu en considérant les trois moyens de transport, c'est-à-dire $P_{\{a,b,r\}}(b) = P_{\{a,b,r\}}(r)$. D'après IIA, on a

$$\frac{P(a, b)}{P(b, a)} = \frac{P_{\{a,b,r\}}(a)}{P_{\{a,b,r\}}(b)}.$$

Ce qui conduit à avoir $P_{\{a,b,r\}}(a) = P_{\{a,b,r\}}(b) = P_{\{a,b,r\}}(r) = \frac{1}{3}$. Or, intuitivement on s'attendrait à ce que la probabilité de choisir le premier bus soit divisée en 2 du fait du second bus ayant des caractéristiques similaires.

(ii) si pour certaines alternatives $a \in T$, $P(a, b) = 0$ pour certaines alternatives $b \in T$, alors

$$P_T(S) = P_{T-\{a\}}(S - \{a\}). \quad (2.2)$$

La seconde partie quant à elle, dit que si certaines alternatives sont toujours choisies en comparaison avec une alternative donnée, alors cette dernière peut être exclue de l'ensemble des alternatives sans pour autant avoir d'impact sur la probabilité du choix des autres alternatives.

Luce montre qu'en cas d'existence d'une fonction d'utilité v positive à une constante près définie sur E , l'équivalence entre (i) de l'axiome et

$$P_S(a) = \frac{v(a)}{\sum_{b \in S} v(b)}. \quad (2.3)$$

En posant $V(b) = \log\{v(b)\}$, c'est-à-dire, $v(b) = \exp\{V(b)\}$, on obtient

$$P_E(a) = \frac{\exp\{V(a)\}}{\sum_{b \in E} \exp\{V(b)\}}, \quad (2.4)$$

formule connue sous le nom du modèle Logit, qui aussi, nous verrons plus tard, peut être obtenue autrement avec la famille des modèles d'Utilité Aléatoire.

Modèle de Tversky

Pour Tversky, l'individu pour opérer son choix, procédera par une élimination aléatoire successive des alternatives selon l'importance qu'il accorde à certaines caractéristiques de ces dernières. Donc sa démarche s'appuie plus sur les caractéristiques de l'alternative que sur l'alternative elle-même. Pour calculer la probabilité de choix d'une alternative, Tversky décrit d'abord chaque alternative par des caractéristiques non décomposables pouvant influencer le choix de l'individu selon que l'alternative les possède ou non, c'est-à-dire 1 si oui et 0 sinon. Pour une caractéristique continue, il propose de définir un seuil pour lequel, on considère que l'alternative possède la caractéristique ou non. Sur les caractéristiques, il définit une fonction d'utilité p . Puis il suppose qu'un individu pour faire son choix parmi ces alternatives, retient une première caractéristique (avec une probabilité reflétant le poids de l'utilité de la caractéristique par rapport aux autres caractéristiques) et de fait, toutes les alternatives ne possédant pas cette caractéristique sont éliminées de son choix. Puis il retient une autre caractéristique pour écarter des alternatives restantes après la première élimination, celles qui ne possèdent pas cette seconde caractéristique. Il procédera ainsi jusqu'à ce qu'il n'y a plus possibilité de nouvelle élimination. A la fin du processus d'élimination :

- ou il reste une seule alternative et dans ce cas, elle est choisie par l'individu, ou bien
- il y a plusieurs alternatives qui restent ; alors chacune d'entre elles serait choisie avec la même probabilité.

Il convient de dire que par cette procédure, toutes caractéristiques communes à toutes les alternatives n'interviennent ni dans le choix ni dans le calcul de la probabilité car ne pouvant permettre de discriminer entre alternatives. Aussi plusieurs séquences d'élimination différentes peuvent amener au choix d'une même alternative. De ce fait, la probabilité de choisir une alternative est la somme des probabilités de toutes ces séquences. Si s est le nombre d'alternatives qui restent après qu'il n'y ait plus de possibilité d'élimination et S_i l'ensemble des alternatives dans $S \in \mathcal{S}$ ayant

en commun la caractéristique i d'utilité p_i pour l'individu, de manière générale, la probabilité que a soit choisie par cet individu est donnée par

$$P_S(a) = \sum_{i=1}^s \frac{p_i}{\sum_{j=1}^s p_j} P_{S_i}(a) \quad (2.5)$$

Pour illustrer la procédure de Tversky, considérons un exemple où un individu devrait opérer un choix parmi 3 alternatives. Cet exemple est celui donné par de Palma [de PALMA and Thisse, 1989]. Notons par E cet ensemble et par a, b, c les trois alternatives décrites par sept caractéristiques de la manière suivante :

- $a = (1, 0, 0, 1, 1, 0, 1)$
- $b = (0, 1, 0, 1, 0, 1, 1)$
- $c = (0, 0, 1, 0, 1, 1, 1)$

Soit p_i , l'utilité de la caractéristique à la $i^{\text{ième}}$ position dans cette description, $i = 1, \dots, 7$.

La caractéristique 7 étant commune à toutes les alternatives, elle n'interviendra pas dans la procédure; donc on se restreindra aux 6 autres caractéristiques pour déterminer la probabilité de choisir chacune des alternatives. On ne donnera que pour une alternative, le procédé étant le même pour les deux autres alternatives. Notons d'abord $W = p_1 + p_2 + p_3 + p_4 + p_5 + p_6$, le poids total des caractéristiques entrant dans le processus du choix et déterminons $P_E(a)$, la probabilité de choisir a .

D'office l'alternative a ne sera jamais choisie si c'est la deuxième ou troisième ou sixième caractéristique qui est retenue par l'individu car elle n'en possède pas; c'est donc des séquences où la probabilité de choisir a est nulle. Mais si l'individu retient la première caractéristique, alors a est choisie car les autres alternatives n'en possèdent pas. La possibilité de choix avec la première caractéristique est obtenue avec la probabilité égale à $\frac{p_1}{W}$, représentant le poids de la caractéristique par rapport à l'ensemble des caractéristiques. Si c'est la quatrième qui est retenue, c est éliminée. En comparant a à b , le choix de a serait fait avec une probabilité de $\frac{p_4}{W}P(a, b)$ et dans le cas où c'est la cinquième caractéristique qui est retenue, ce sera une comparaison entre a et c . Dans ce cas a serait choisie avec une probabilité de $\frac{p_5}{W}P(a, c)$. Les trois cas possibles de choix de a que nous venons d'évoquer étant exclusifs, alors

$$P_E(a) = \frac{p_1}{W} + \frac{p_4}{W}P(a, b) + \frac{p_5}{W}P(a, c). \quad (2.6)$$

Il faut noter que $P(a, b)$ et $P(a, c)$ représentent les poids de caractéristiques appartenant à a et n'appartenant pas à l'autre alternative par rapport aux caractéristiques entrant dans leur comparaison, i.e

$$P(a, b) = \frac{p_1 + p_5}{p_1 + p_2 + p_5 + p_6},$$

$$P(a, c) = \frac{p_1 + p_4}{p_1 + p_3 + p_4 + p_5}.$$

Ce modèle a l'avantage d'être une généralisation du modèle de Luce. En effet quand $S_i\{a\}$, $P_{S_i}(a) = 1$, l'équation (2.5) devient

$$P_S(a) = \frac{p_i}{\sum_{j=1}^s p_j} \quad (2.7)$$

qui es la formule du modèle de Luce.

Un point faible de ce modèle est le difficile calcul de la probabilité de choix d'une alternative. Rien qu'avec ces trois alternatives, on constate que le calcul de la probabilité d'une alternative est laborieux. La difficile énumération de toutes les possibles séquences rend compliquée la mise en œuvre de ce modèle. Un autre point faible est qu'une caractéristiques a un même poids pour toutes les alternatives. Ce qui peut sembler irréaliste car en fonction de la présence ou non de certaines caractéristiques, un individu peut plus ou moins accorder d'importance à d'autres caractéristiques d'une alternative.

2.1.2 Les modèles d'Utilité Aléatoire (RUMs)

Pour un aperçu plus approfondi sur les modèles d'Utilité Aléatoire, les démonstrations des formules et les méthodes d'estimation, [Train, 2009] est une bonne référence. D'ailleurs cette sous-section s'en inspire pour beaucoup en ce qui concerne la présentation des modèles.

Le concept de base des RUMs

Soient J alternatives mutuellement exclusives parmi lesquelles un individu devrait choisir une et une seule alternative. Contrairement aux modèles de choix à règle de décision aléatoire, les modèles d'Utilité Aléatoire supposent que la règle de choix est déterministe et que l'individu n choisirait l'alternative j qui lui procurerait le plus d'utilité. En d'autres termes, si U_{nj} est l'utilité qu'obtiendrait l'individu n avec l'alternative j ($j = 1, \dots, J$), alors il choisirait l'alternative j telle que $U_{nj} > U_{ni}, \forall i \neq j$. Sauf que si l'individu n connaît l'utilité U_{nj} qu'il obtient de l'alternative j , le chercheur ne connaît qu'une partie de cette utilité; partie décrite par des variables observables que nous nommons x_{nj} . L'utilité U_{nj} peut donc être vue par le chercheur comme la somme de deux parties; une, $\beta'x_{nj}$ (β vecteur de paramètres) que nous notons $V_{nj}(\beta)$, appelée utilité représentative et décrite par les variables observables x_{nj} et l'autre, notée ϵ_{nj} , inconnue au chercheur. On écrit alors

$$U_{nj} = V_{nj}(\beta) + \epsilon_{nj}. \quad (2.8)$$

La partie inconnue ϵ_{nj} est traitée par le chercheur comme étant aléatoire et lui assigne une distribution; ce qui permet de faire des calculs probabilistes sur l'éventuel choix de n'importe quelle alternative j . Ainsi la probabilité que l'alternative j soit choisie par l'individu n , est la probabilité, P_{nj} , que cette alternative ait une utilité plus

grande que celle de chacune des autres alternatives, i.e,

$$\begin{aligned}
P_{nj} &= \text{Prob}(U_{nj} > U_{ni}, \forall i \neq j) \\
&= \text{Prob}(V_{nj} + \epsilon_{nj} > V_{ni} + \epsilon_{ni}, \forall i \neq j) \\
&= \text{Prob}(\epsilon_{nj} - \epsilon_{ni} > V_{ni} - V_{nj}, \forall i \neq j)
\end{aligned} \tag{2.9}$$

En fonction de la spécification de la distribution $f(\epsilon_{n1}, \dots, \epsilon_{nJ})$, spécification qui dépend plus du choix de la modélisation que de l'aléa, différents modèles sont obtenus. Nous présenterons quatre de ces modèles (les plus connus) dans les paragraphes qui suivent.

Remarque Le fait que la probabilité P_{nj} soit basée sur la différence des utilités a deux implications : en augmentant toutes les utilités d'une même constante ou en les multipliant par un même scalaire positif, la probabilité de choix d'une alternative ne change pas ; ce qui peut conduire aussi à des problèmes d'identification de certains paramètres.

Modèle Logit

Comme dit dans la sous-section 2.1.1, le Logit a d'abord été développé par Luce [Luce, 1959] avant même les travaux de Marschak [Marschak, 1960] qui ont consacré l'appellation des modèles d'Utilité Aléatoire plus connus sous l'acronyme anglais de RUMs (Random Utility Models). Pour les RUMs, le Logit est obtenu sous l'hypothèse que les $\epsilon_{nj}/\sigma, j = 1, \dots, J$ sont indépendamment et identiquement distribués selon la loi de Gumbel, i.e,

$$f(\epsilon_{nj}/\sigma) = e^{-\epsilon_{nj}/\sigma} e^{-e^{-\epsilon_{nj}/\sigma}},$$

où σ^2 est la variance commune des ϵ_{nj} .

La probabilité de choix d'une alternative (2.9) est donnée par

$$P_{nj} = \frac{e^{V_{nj}/\sigma}}{\sum_{i=1}^J e^{V_{ni}/\sigma}}. \tag{2.10}$$

Comme on le voit, ce modèle est de forme simple et facile à mettre en œuvre. Il possède des propriétés intéressantes comme par exemple celles que P_{nj} est comprise entre 0 et 1 et de forme sigmoïde ou que $\sum_j P_{nj} = 1$; des propriétés qui rendent facile l'interprétation des résultats. Cette simplicité de forme et ses propriétés ont probablement fait qu'il est le plus connu et le plus utilisé dans la modélisation des données de choix discret. Seulement, ce modèle possède un certain nombre de limitations qui font que dans certaines situations son utilisation soit critiquable.

D'abord, il est obtenu sous l'hypothèse d'indépendance des parties inconnues ϵ_{nj} .

Le ratio des probabilités de choix

$$\frac{P_{nj}}{P_{ni}} = \frac{\frac{e^{\beta x_{nj}}}{\sum_{i=1}^J e^{\beta x_{ni}}}}{e^{\beta x_{nj}}} = \frac{e^{\beta x_{nj}}}{\sum_{i=1}^J e^{\beta x_{ni}}}$$

est indépendant des autres alternatives. Autrement dit, la présence d'autres alternatives n'a pas d'impact sur ce ratio ; c'est donc dire qu'on peut ajouter ou retirer une autre alternative sans changer ce ratio. Pourtant dans certaines situations de choix, l'ajout ou le retrait d'une alternative est sensé modifier ce rapport. C'est le cas du fameux paradoxe du bus bleu-bus rouge que nous avons vu précédemment (2.1.1). Puis ce modèle ne peut s'utiliser dans le cas où il y a une variation de goût aléatoire chez les individus. En effet, les individus ne placent pas la même importance pour une caractéristique : par exemple un individu avec un fort revenu est moins sensible au coût de transport qu'un autre ayant un faible revenu. Si cette variation est aléatoire, alors elle ne peut être captée par le Logit. Une troisième limitation concerne une dépendance temporelle de la partie inconnue, c'est à dire quand une série d'observations est faite sur un individu dans le temps,

$$U_{njt} = V_{njt} + \epsilon_{njt}$$

et que $\epsilon_{njt}, t = 1, \dots, T$ sont corrélées, le Logit ne peut être utilisé. Pour essayer de lever ces limitations, divers autres modèles ont été développés.

Modèle Nested-Logit

Pour répondre à la limitation du Logit concernant la dépendance entre ϵ_{nj} , les modèles GEV (General Extreme Value), plus généraux que ce dernier, ont été développés. Le plus utilisé est le Nested-Logit [Ben-Akiva, 1972]. Pour ce modèle, l'ensemble des J alternatives est partitionné en K sous-ensembles B_k appelés nœuds (Nests en Anglais), de telles sortes que si deux alternatives j et m appartiennent toutes à un même nœud, ϵ_{nj} et ϵ_{nm} sont corrélées et si elles appartiennent à deux sous-ensembles distincts, alors ϵ_{nj} et ϵ_{nm} sont indépendantes. La distribution jointe cumulée des ϵ_{nj} est donnée par

$$F(\epsilon_{n1}, \dots, \epsilon_{nJ}) = \exp \left\{ - \sum_{k=1}^K \left(\sum_{i \in B_k} e^{-\epsilon_{ni}/\lambda_k} \right)^{\lambda_k} \right\}, \quad (2.11)$$

et la probabilité de choix associée est

$$P_{nj} = \frac{e^{V_{nj}/\lambda_k} \left(\sum_{m \in B_k} e^{V_{nm}/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{l=1}^K \left(\sum_{i \in B_l} e^{V_{ni}/\lambda_l} \right)^{\lambda_l}}. \quad (2.12)$$

Le paramètre λ_k mesure le degré de corrélation entre les parties inconnues des alternatives appartenant au nœud B_k . Plus petit est ce paramètre, plus corrélées sont les parties inconnues des alternatives appartenant à ce nœud. Avec $\lambda_k = 1$ pour tout $k = 1, \dots, K$, on retrouve la formule de Logit.

Modèle Probit

C'est le plus ancien des modèles d'Utilité Aléatoire. Il a été développé par Thurstone [Thurstone, 1927] pour le cas binaire dans le cadre d'une étude sur le stimuli. Il est obtenu dans le cas où la distribution jointe des ϵ_{nj} est une loi normale multivariée, i.e,

$$f(\epsilon_{n1}, \dots, \epsilon_{nJ}) = \frac{1}{(2\pi)^{-J/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} \epsilon_n \Sigma^{-1} \epsilon_n^T},$$

où $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nJ})$, ϵ_n^T son transposé et Σ la matrice de covariance des ϵ_{nj} . La probabilité de choix d'une alternative est donnée par

$$P_{nj} = \int_{\epsilon} I(\epsilon_{ni} - \epsilon_{nj} < V_{nj} - V_{ni}, \forall i \neq j) f(\epsilon_{n1}, \dots, \epsilon_{nJ}) d\epsilon_{n1} \dots d\epsilon_{nJ}. \quad (2.13)$$

Ce modèle permet de gérer toutes les limitations du Logit évoquées plus haut. Son seul problème est que la forme de probabilité de choix n'est pas simple ; ce qui peut engendrer d'abord un processus d'identification des paramètres de la matrice Σ assez laborieux et/ou amène à faire recours aux simulations pour estimer les paramètres.

Mixed-Logit

Ce modèle a probablement été utilisé en premier par [Boyd and Mellman, 1980] ou [Cardell and Dunbar, 1980] pour la modélisation des données du marché d'automobile. Il est obtenu juste en moyennant la probabilité de choix du Logit sur le paramètre β . Autrement dit, à β est attribué une densité $g(\beta)$ et P_{nj} est définie par

$$P_{nj} = \int_{\beta} \frac{e^{V_{nj}(\beta)}}{\sum_j e^{V_{nj}(\beta)}} g(\beta) d\beta, \quad (2.14)$$

g peut-être n'importe quelle distribution. Le modèle est équivalent au Logit quand g est une distribution de Dirac sur β . Il est le plus flexible de tous les RUMs [McFadden et al., 2000]. Bien que la forme de la probabilité peut être complexe, la simulation est plus simple à mettre en œuvre que pour le modèle Probit. Comme Probit, il permet de gérer toutes les limitations du Logit.

2.2 Géostatistique : le processus gaussien

Les données d'admission sont spatialement réparties avec possibilité de dépendance entre elles. Une approche pour leur modélisation peut relever du domaine de

la géostatistique qui est l'étude des variables régionalisées. Et un modèle souvent utilisé en géostatistique est le processus gaussien. Dans cette section, nous présentons ce processus avec quelques exemples de fonctions de covariance connues pour traduire la nature de la dépendance spatiale entre les différentes localisations des données. D'autres processus comme le champ de Markov ou le processus de Poisson existent et auraient pu être utilisés mais nous n'en faisons pas cas car dans notre travail (chapitres 2 et 3), nous n'avons utilisé que le champ gaussien. Outre qu'il s'utilise pour des données continues transformées (pour des exemples de transformation de données, voir [Box and Cox, 1964]) ou non, il peut aussi s'utiliser très bien pour des données discrètes à travers des modèles hiérarchiques. Par exemple pour des données suivant une loi de Poisson de moyenne λ ou une loi binomiale de paramètre p , il peut s'utiliser sur $\log(\lambda)$ ou sur θ avec $p = \frac{1}{1 + \exp(-\theta)}$. C'est un modèle, même en dehors de toute justification physique, qui peut être utilisé pour sa maniabilité et son extrême flexibilité pour ajuster les données.

Pour une vue plus générale sur le sujet, le livre de de Noël Cressie, *Statistics for spatial data*, sorti en 1993 (dont une ré-édition est [Cressie, 2015]) pourrait être une très bonne référence pour un lecteur intéressé.

2.2.1 Champ gaussien

On appelle processus stochastique $S = (S_t)_{t \in T}$, une famille de variables aléatoires S_t indexée par un ensemble T . Lorsque $T = \mathbb{R}^2$, on parle de champ stochastique et dans le cas échant, nous utilisons x au lieu de t et $S(x)$ au lieu de S_t , x désignant la localisation de la réalisation de la variable aléatoire. Le champ est dit gaussien, si le vecteur $\{S(x_1), \dots, S(x_n)\}$ suit une loi normale multivariée de vecteur-moyenne μ et de matrice de covariance Σ . Le champ est spécifié par la donnée des fonctions $\mu(x) = \mathbb{E}\{S(x)\}$ (espérance mathématique du champ au point x) et $\gamma(x, x') = Cov\{S(x), S(x')\}$ qui sont telles que μ est le transposé de $\{\mu(x_1), \dots, \mu(x_n)\}$ et les éléments de la matrice Σ , $\Sigma_{n,ij} = \gamma(x_i, x_j)$.

On parle de champ gaussien stationnaire quand $\mu(x)$ est constante et $\gamma(x, x') = \gamma(x, x + h)$, ne dépend que de $x' - x = h$; alors on écrit $\gamma(h)$ au lieu de $\gamma(x, x')$. Il est isotrope quand $\gamma(h)$ ne dépend que de la distance $\|h\|$.

2.2.2 Exemples de fonctions de covariance isotropes

Comme dit ci-dessus, un champ gaussien S est complètement déterminé par la donnée du vecteur-moyenne transposé de $\{\mu(x_1), \mu(x_2), \dots, \mu(x_n)\}$ et de la matrice de covariance Σ d'éléments $\Sigma_{ij} = \gamma(x_i, x_j)$ pour tout vecteur $\{S(x_1), S(x_2), \dots, S(x_n)\}$. Si le vecteur-moyenne peut être n'importe quelle fonction, il n'en est pas de même pour la matrice de covariance, donc pour la fonction de covariance. En effet, $Y = \sum_{i=1}^n a_i S(x_i)$ où $a_i \in R$, suit une loi normale de moyenne $\mu_Y = \sum_{i=1}^n \mu(x_i)$ et de variance $\sigma_Y^2 = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \Sigma_{ij} = a' \Sigma a$, avec $a = (a_1, a_2, \dots, a_n)$. Comme la va-

riance est positive, alors $\sum_{i=1}^n \sum_{j=1}^n a_i a_j \Sigma_{ij} = a' \Sigma a \geq 0$ pour tous (x_1, x_2, \dots, x_n) et (a_1, a_2, \dots, a_n) ; contraignant ainsi Σ à être une matrice définie positive et la fonction $\gamma(\cdot)$ correspondante donc à être une fonction définie positive. Cette définie positivité peut être vérifiée à partir des critères de Schlather [Schlather, 1999] qui s'est appuyé sur les travaux de Gneiting [Gneiting, 1997]. Et pour les fonctions de covariance isotropes, elles peuvent être construites ou vérifiées grâce au théorème de Schoenberg [Schoenberg, 1938] qui dit qu'une fonction $\gamma(\cdot)$ est une fonction de covariance isotrope dans R^d si et seulement si elle vérifie

$$\gamma(h) = \int_0^\infty \exp\{-(\|h\|/t)^2\} d\nu(t),$$

avec ν , mesure positive bornée définie sur R^+ .

Dans la pratique, pour s'assurer que quels que soient (x_1, x_2, \dots, x_n) , toute combinaison $\sum_{i=1}^n a_i S(x_i)$ a une variance positive, le choix d'une fonction de covariance s'opère dans des familles de fonctions bien connues; familles de fonctions qui peuvent couvrir pas mal de cas réels surtout en ce qui concerne la famille de Matérn.

Nous citons pour exemples de ces fonctions (donc vérifiant le théorème de Schoenberg) :

-famille dite de fonctions de covariance stables

Elles sont définies par

$$\gamma(h) = \tau \exp\{-(\|h\|/\omega)^\nu\},$$

avec ω, τ strictement positifs et $0 < \nu \leq 2$.

Elles traduisent comme dans beaucoup de phénomènes, une décroissance monotone plus ou moins rapide selon la valeur de ν au fur et à mesure que les points sont distants. Avec $\nu = 1$, on a une décroissance exponentielle et avec $\nu = 2$, une fonction de covariance dite gaussienne du fait de sa forme analytique.

-famille de Matérn

Flexibles grâce au paramètre ν , les fonctions de covariance de Matérn [Matérn, 1960] sont définies par

$$\gamma_\nu(h) = \tau \{2^{\nu-1} \Gamma(\nu)\}^{-1} (\|h\|/\omega)^\nu \mathbf{K}_\nu(\|h\|/\omega),$$

où τ, ω et ν strictement positifs et $\mathbf{K}_\nu(\cdot)$ la fonction de Bessel modifiée d'ordre ν . Pour $\nu = \frac{1}{2}$, on a $\gamma_{1/2}(h) = \tau \exp\{-\|h\|/\omega\}$, la fonction de covariance exponentielle et lorsque $\nu \rightarrow +\infty$, $\gamma_{+\infty}(h) \rightarrow \tau \exp\{-(\|h\|/\omega)^2\}$, la fonction de covariance dite gaussienne, deux toutes comme on l'a vu, appartenant à la famille de fonctions de covariance dites stables. La famille de Matérn présente plus de régularité quant à la

différentiabilité par rapport à la famille des fonctions de covariance précédemment citée. Et tout comme la famille exponentielle, la corrélation diminue au fur et à mesure que les localisations sont distantes.

-autres familles

Il y a par exemple la famille sphérique, utilisée surtout par les géologues, définie par

$$\gamma(h) = \begin{cases} \tau \left\{ 1 - \frac{3}{2}(\|h\|/\omega) + \frac{1}{2}(\|h\|/\omega)^3 \right\}, & 0 \leq \|h\| \leq \omega \\ 0, & \omega > \|h\|. \end{cases}$$

Il faut noter qu'avec cette famille, le paramètre ω dicte la distance au delà de laquelle, il n'existe plus de corrélation entre deux localisations.

Il y a aussi la famille non-monotone définie par

$$\gamma(h) = \tau(\|h\|/\omega)^{-1} \sin(\|h\|/\omega).$$

pouvant exprimer une sorte de périodicité de la corrélation spatiale.

2.3 Éléments de la Statistique Bayésienne

Une des approches en statistique en vue d'une analyse (description ou prédiction) est de modéliser le comportement d'un phénomène en le supposant régi par une variable aléatoire. La distribution des observations \mathbf{x} de cette variable est représentée par une fonction de densité $f(\mathbf{x} \mid \theta)$ appelée vraisemblance. θ est le paramètre gouvernant la distribution et est de dimension finie. Mais dans la pratique, il n'est pas toujours connu ; donc il faudra l'estimer.

2.3.1 Modélisation bayésienne

Étant donné

$$f(\mathbf{x} \mid \theta), \quad \theta \in \Theta,$$

plusieurs méthodes existent pour estimer θ . Par exemple, la méthode consistant à maximiser la vraisemblance peut être utilisée ou bien aussi la méthode dite des moments ; des méthodes basées seulement sur les informations apportées par les observations pour l'estimation du paramètre. Pourtant, il est possible d'avoir des intuitions ou des informations sur θ sans même faire recours aux observations disponibles et qui une fois utilisées, ces informations peuvent bien améliorer l'estimation du paramètre. Le paradigme bayésien suppose que θ est aléatoire et qu'une loi de probabilité π sur Θ , dite loi *a priori*, peut-être utilisée pour traduire ces informations ou intuitions.

De ce fait, un modèle paramétrique bayésien est donné par :

$$\begin{aligned} \mathbf{x} &\sim f(\mathbf{x} \mid \theta), & \theta &\in \Theta & \text{et} \\ \theta &\sim \pi(\theta) \end{aligned} \tag{2.15}$$

Les observations, dans cette approche, se trouvent comme un apport supplémentaire d'informations conduisant à une actualisation des connaissances sur θ .

La formule de Bayes [Bayes, 1763], dont une version

$$\pi(\theta | \mathbf{x}) = \frac{f(\mathbf{x} | \theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x} | \theta)\pi(\theta) d\theta} \propto f(\mathbf{x} | \theta)\pi(\theta), \quad (2.16)$$

donne cette actualisation par la distribution $\pi(\theta | \mathbf{x})$, dite distribution *a posteriori*, en incorporant les informations apportées par les observations. Cette distribution est la distribution sur laquelle se fait, en statistique bayésienne, toute inférence sur le paramètre θ .

2.3.2 Inférence bayésienne

La loi *a posteriori*, $\pi(\theta | \mathbf{x})$, peut être résumée à travers certaines valeurs de θ dites **optimales** (au sens du critère utilisé pour l'estimation), donnant donc lieu à des estimations ponctuelles de θ . Mais aussi, elle permet de donner des intervalles de crédibilité de θ , notion somme toute naturelle, car θ est considéré comme aléatoire.

Pour déterminer un estimateur ponctuel θ^* optimal du paramètre θ , des fonctions de coût ou de perte, $\mathcal{L}(\theta^*, \theta)$, peuvent être utilisées. Il s'agit des fonctions quantifiant le coût à payer ou la perte à subir en utilisant θ^* au lieu du vrai paramètre θ . Même si on peut définir une fonction de perte spécifique au besoin (mais pouvant rendre encore plus difficile l'inférence), les fonctions les plus couramment utilisées sont :

– la fonction Perte 0-1 :

$$\mathcal{L}(\theta^*, \theta) = \begin{cases} 0, & \text{si } \|\theta^* - \theta\| \leq \epsilon \\ 1, & \text{si } \|\theta^* - \theta\| \geq \epsilon \end{cases} \quad (2.17)$$

où ϵ donne la précision sur l'erreur admise.

– la fonction Perte absolue :

$$\mathcal{L}(\theta^*, \theta) = \|\theta^* - \theta\| \quad (2.18)$$

– la fonction Perte quadratique :

$$\mathcal{L}(\theta^*, \theta) = (\theta^* - \theta)^2 \quad (2.19)$$

La valeur optimale selon le critère est celle qui minimise la fonction. Par exemple, la fonction Perte 0-1 est minimisée par le mode de la distribution *a posteriori*. La fonction Perte Absolue est minimisée par la médiane. Quant à la fonction Perte quadratique, elle est minimisée par l'espérance de la loi *a posteriori*, c'est-à-dire $\theta^* = \mathbb{E}^\pi(\theta | \mathbf{x})$; qui est donc la valeur optimale au sens de ce critère. C'est cette valeur que nous avons utilisée pour résumer les différentes lois *a posteriori* dans nos travaux de recherche afin de faire la comparaison de l'attractivité des hôpitaux.

Comme évoqué tantôt, la loi a posteriori permet de construire ou de déterminer des intervalles de crédibilité sous forme des régions à plus forte densité a posteriori. Associés à une probabilité α de contenir θ , sachant les observations \mathbf{x} , ces intervalles sont de la forme :

$$C_\alpha(\mathbf{x}) = \{\theta : \pi(\theta | \mathbf{x}) \geq k\} \quad (2.20)$$

avec k tel que

$$\int_{C_\alpha(\mathbf{x})} \pi(\theta | \mathbf{x}) d\theta = \alpha. \quad (2.21)$$

Ces intervalles renseignent par exemple sur la précision d'une estimation.

2.3.3 Choix de modèles

Il arrive qu'au moins deux modèles soient concurrents pour la modélisation d'un même ensemble d'observations \mathbf{x} . La question qui se pose est celle de choix du modèle qui semble le mieux adapté. Plusieurs méthodes existent pour opérer cette sélection dont le facteur de Bayes ou le Deviance Information Criterion (DIC) [Spiegelhalter et al., 2002]. Pour deux modèles \mathcal{M}_i , $i = 1, 2$ tels que

$$\begin{aligned} \mathcal{M}_i : f_i(x | \theta_i), \theta_i \in \Theta_i, \\ \theta_i \sim \pi_i(\theta_i), \end{aligned} \quad (2.22)$$

le facteur de Bayes est défini par

$$B_{12}(x) = \frac{\mathbb{P}(\mathcal{M}_1 | x)}{\mathbb{P}(\mathcal{M}_2 | x)} / \frac{\mathbb{P}(\mathcal{M}_1)}{\mathbb{P}(\mathcal{M}_2)} \quad (2.23)$$

où $\mathbb{P}(\mathcal{M}_i)$ sont des lois a priori sur les modèles \mathcal{M}_i et $\mathbb{P}(\mathcal{M}_i | x)$ telles que

$$\begin{aligned} \mathbb{P}(\mathcal{M}_1) + \mathbb{P}(\mathcal{M}_2) = 1 \quad \text{et} \\ \mathbb{P}(\mathcal{M}_i | x) \propto \mathbb{P}(\mathcal{M}_i) \int_{\Theta_i} f_i(x | \theta_i) \pi_i(\theta_i) d\theta_i. \end{aligned} \quad (2.24)$$

Si B_{12} est supérieur à 1, on dira que le modèle \mathcal{M}_1 est meilleur que \mathcal{M}_2 , sinon, le contraire.

Le DIC, plus utilisé en cas d'échantillons obtenus par MCMC, est défini par

$$DIC_i = 2\mathbb{E}_{\theta_i|x}(D_i) - D_i[\mathbb{E}_{\theta_i|x}(\theta_i)]. \quad (2.25)$$

D_i est la déviance définie par

$$D_i = -2[l(x | \hat{\theta}_i) - l(x | \theta_{sature})], \quad (2.26)$$

$\hat{\theta}_i$ la valeur estimée, θ_{sature} , la valeur pour le modèle de référence (saturé) et l la fonction de log-vraisemblance du modèle. Plus petit est le DIC, plus bon c'est.

2.3.4 Remarques

Remarque 1 : Il est à noter que le début d'une inférence bayésienne commence d'abord par le choix de la loi a priori. Deux lois a priori différentes peuvent conduire à des résultats complètement différents sur un même modèle d'observations surtout en ce qui concernent l'intervalle de crédibilité même si avec l'estimation ponctuelle, le poids de la loi a priori tend à disparaître asymptotiquement, c'est-à-dire avec un très large échantillon d'observations. Donc le choix de la loi a priori est crucial et il convient donc de lui accorder un grand soin. Plusieurs manières existent pour faire ce choix dont :

Loi conjuguée :

Il s'agit de choisir $\pi(\theta)$ appartenant à une famille de distributions \mathcal{F} sur Θ telle que $\pi(\theta | x)$ appartienne aussi à cette famille \mathcal{F} .

Loi d'entropie maximale Si un certain nombre (K fini) de caractéristiques de la loi a priori sont connues et telles que

$$\mathbb{E}^\pi[g_k(\theta)] = w_k, \quad k = 1, \dots, K, \quad (2.27)$$

alors π est choisie de telle sorte, pour un nombre fini n , que l'expression

$$E_\pi = - \sum_i^n \pi(\theta_i) \log\{\pi(\theta_i)\} \quad (2.28)$$

appelée entropie, soit maximale.

Loi non informative

Il s'agit des lois uniformes ou des extensions de ces dernières. Elles ont peu de poids par rapport à la fonction de vraisemblance sur les résultats de l'inférence.

En ce qui nous concerne, nous avons préféré cette dernière approche en utilisant, ou des lois uniformes, ou des lois très plates (proche des lois uniformes) pour ne pas être informatives sauf mention contraire.

Remarque 2 : La loi a posteriori peut être d'une forme complexe au point de ne pas être connue analytiquement. Alors pour les estimations ponctuelles ou détermination des intervalles de crédibilité, on fera recours aux échantillons distribués selon cette loi, échantillons qui peuvent être obtenus par méthodes de Monte Carlo par Chaînes de Markov (MCMC).

2.4 Méthodes de Monte Carlo par Chaînes de Markov

2.4.1 Définition

On appelle méthodes de Monte Carlo, en référence aux jeux de hasard pratiqués à Monte Carlo, toute méthode visant à calculer une intégrale ou toute autre quantité

numérique en utilisant des procédés aléatoires, c'est-à-dire en utilisant des résultats de la théorie des probabilités. Par exemple, la quantité

$$I = \int_0^1 g(x) dx$$

serait approximée par

$$\frac{1}{n} \sum_{i=1}^n g(x_i),$$

où (x_1, \dots, x_n) sont des réalisations indépendantes de la loi uniforme sur l'intervalle $(0, 1)$. Cette approximation se justifie par la loi des grands nombres qui dit que pour N suffisamment grand,

$$\frac{1}{N} \sum_{i=1}^n g(x_i)$$

converge presque sûrement vers I .

Plus généralement, si $f(x)$ est telle que

$$\begin{aligned} f(x) &> 0, \quad \forall x \in \mathcal{X} \quad \text{et} \\ \int_{\mathcal{X}} f(x) dx &= 1, \end{aligned} \tag{2.29}$$

alors la quantité

$$I = \int_{\mathcal{X}} g(x) f(x) dx = \mathbb{E}^f[g(x)]$$

peut être estimée par

$$\hat{g}_N = \frac{1}{N} \sum_{i=1}^N g(x_i)$$

avec x_i des réalisations indépendantes de la loi f . Et lorsque g est de carré intégrable par rapport à $f(x)$, la variance de l'estimation est

$$\text{var}(\hat{g}_n) = \frac{1}{n} \left[\int_{\mathcal{X}} g^2(x) f(x) dx - \{\mathbb{E}^f[g(x)]\}^2 \right],$$

qui peut être évaluée par

$$v_N = \frac{1}{n} \sum_{i=1}^n g^2(x_i) - \frac{\hat{g}_n^2}{n^2}$$

et permettant de contrôler la qualité de l'estimation de I par \hat{g}_N . Quand N est suffisamment grand, grâce au Théorème Central Limite, on peut supposer que

$$\frac{\hat{g}_N - \mathbb{E}^f\{g(x)\}}{\sqrt{v_N}} \sim \mathcal{N}(0, 1).$$

Ce qui permet la construction des intervalles de crédibilité sur l'approximation.

Mais, le problème est qu'il n'est pas toujours possible d'avoir directement des réalisations de f , surtout quand f est une distribution a posteriori qui peut être de forme très complexe dans l'inférence bayésienne. Il faudra faire recours à des algorithmes. Et on appelle méthodes de Monte Carlo par Chaînes de Markov (MCMC), celles basées sur des algorithmes qui permettent de produire un échantillon selon la distribution f , en utilisant des propriétés d'une chaîne de Markov.

Le principe de base de ces algorithmes est, à partir d'un noyau de transition de loi stationnaire f et d'une valeur initiale arbitraire $x^{(0)}$, de générer une chaîne de Markov $x^{(t)}$, convergeant vers la distribution f . Il faudra donc un certain nombre d'itérations suffisamment grand avant d'avoir un échantillon distribué selon la loi f . C'est des techniques qui peuvent être onéreuses en temps de calcul. Dans ce chapitre nous présentons deux de ces algorithmes que nous avons utilisés, à savoir l'échantillonnage de Gibbs et l'algorithme de Hastings-Metropolis.

2.4.2 Échantillonnage de Gibbs

L'échantillonnage de Gibbs [Gelfand and Smith, 1990] est un algorithme MCMC qui permet de simuler selon une loi d'intérêt à partir de ses lois conditionnelles. Ce qui suppose que nécessairement la loi en question est multidimensionnelle, quitte à ajouter des composantes artificielles. Il suppose aussi une connaissance préalable de certaines propriétés probabilistes ou analytiques de la loi car ses lois conditionnelles doivent être simulables. Ainsi pour simuler selon une loi d'intérêt, $f(\mathbf{x})$ avec $\mathbf{x} = (x_1, \dots, x_p)$ et $p > 1$, dont les lois conditionnelles f_1, \dots, f_p correspondantes aux composantes de \mathbf{x} sont simulables, l'algorithme d'échantillonnage de Gibbs associé a pour transition de $\mathbf{x}^{(t)}$ à $\mathbf{x}^{(t+1)}$:

Simuler

1. $x_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \dots, x_p^{(t)});$
2. $x_2^{(t+1)} \sim f_2(x_2 | x_1^{(t+1)}, x_3^{(t)}, \dots, x_p^{(t)});$
- ...
- p. $x_p^{(t+1)} \sim f_p(x_p | x_1^{(t+1)}, \dots, x_{p-1}^{(t+1)}).$

Il convient de remarquer que toutes les valeurs simulées sont acceptées comme dans le cas d'une simulation à partir d'une loi d'intérêt directement simulable.

2.4.3 Algorithme Hastings-Metropolis

Contrairement à l'algorithme d'échantillonnage de Gibbs, l'algorithme de Hastings-Metropolis [Gelman et al., 2004] n'exige pas une connaissance poussée de la loi d'intérêt $f(x)$ et cette loi n'est pas forcément multidimensionnelle. Il repose sur l'utilisation d'une loi conditionnelle $q(y | x)$ qui est simulable rapidement et est soit analytiquement connue (à une constante près) ou est symétrique, c'est-à-dire $q(x | y) = q(y | x)$. La loi q est appelée loi instrumentale ou de proposition. Étant

donné une loi $f(x)$ et une loi conditionnelle $q(y | x)$, la chaîne $(x^{(t)})$ est générée de la manière suivante :

1. Générer $y_t \sim q(y|x^{(t)})$.

2. Prendre

$$x^{(t+1)} = \begin{cases} y_t & \text{avec proba } \rho(x^{(t)}, y_t), \\ x^{(t)} & \text{avec proba } 1 - \rho(x^{(t)}, y_t) \end{cases}$$

où

$$\rho(x^{(t)}, y_t) = \min \left\{ \frac{f(y_t) q(x^{(t)} | y_t)}{f(x^{(t)}) q(y_t | x^{(t)})}, 1 \right\}.$$

Notons qu'avec cet algorithme, toutes les valeurs simulées ne sont pas systématiquement acceptées contrairement à l'algorithme d'échantillonnage de Gibbs. Elles ne le sont systématiquement que lorsque $\frac{f(y_t)}{q(y_t|x^{(t)})} > \frac{f(y_{t-1})}{q(y_{t-1}|x^{(t-1)})}$. Dans le cas où q est symétrique, l'acceptation est gouvernée par le rapport $f(y_t)/f(x^{(t)})$. La probabilité $\rho(x^{(t)}, y_t)$ n'est définie que si $f(x^{(t)})$ est non nul. Il convient donc de choisir $x^{(0)}$ (valeur initiale), et c'est suffisant, tel que $f(x^{(0)})$ soit strictement supérieur à zéro. Aussi le support de la loi de proposition doit recouvrir celui de la loi d'intérêt pour que la chaîne puisse explorer toutes les zones de cette dernière.

2.4.4 Remarques

Remarque 1 : L'algorithme de Hastings-Metropolis est plus générique en ce sens qu'il permet de simuler selon pratiquement n'importe quelle loi sauf que le manque de lien entre la loi de proposition et la loi d'intérêt peut être néfaste pour la convergence de la chaîne. L'algorithme d'échantillonnage de Gibbs peut être vu comme une combinaison d'algorithmes de Hastings-Metropolis où la loi de proposition pour chaque loi conditionnelle est la loi conditionnelle elle-même. L'échantillonnage de Gibbs exploite bien les propriétés de la loi mais il entraîne des choix très limités des lois de proposition.

Remarque 2 : Dans nos travaux, il arrive que même les lois conditionnelles devant être utilisées dans l'échantillonnage de Gibbs ne sont pas explicitement connues ; ce qui fait que pour chaque étape de l'échantillonnage de Gibbs, nous utilisons l'algorithme de Hastings-Metropolis pour mettre à jour les paramètres.

Remarque 3 : Comme dit si haut, il faudra un grand nombre d'itérations, appelé temps de chauffe de la chaîne, avant d'atteindre la convergence. Pour le contrôle de cette convergence, nous avons utilisé le graphique de la chaîne et la stabilité de la moyenne a posteriori assurée par le **Théorème ergodique** mais aussi pour un des papiers nous avons utilisé R.hat statistics [Aho, 2015] qui permet de savoir si les chaînes proviennent d'une même distribution. Ce qui nous amène à faire plusieurs chaînes pour vérifier cette convergence.

CHAPITRE 3

Effet de la distance sur le nombre d'admission de patients au niveau d'un hôpital

Le but de l'étude est de mesurer l'attractivité des hôpitaux et subséquemment les classer du plus attractif au moins attractif. Le nombre moyen de patients admis en fonction de la distance qui sépare les lieux de résidence des patients à l'hôpital peut être révélateur de l'attractivité de l'hôpital dans une zone où des centres de soins de santé sont disponibles à des distances relativement courtes des lieux de résidence des patients. En effet, plus un hôpital est attractif, plus la distance compte moins pour s'y rendre. Donc un hôpital qui a plus d'admissions au fur à mesure qu'on s'en éloigne est plus attractif que celui qui en a moins. Pour ce faire, nous disposons, grâce au *PMSI*, d'informations telles que le lieu de résidence du patient, la pathologie pour laquelle le patient est admis et dans quel hôpital il l'est. Grâce à ces informations, un regroupement de patients par localité et par hôpital pour une pathologie donnée peut être effectué. Ces données d'admission peuvent être vues comme des données de comptage spatialement réparties qui peuvent être dépendantes comme il est souvent le cas pour les données de santé publique. Il est aussi possible dans une zone prédéfinie pour l'étude, d'avoir des localités avec zéro admis dans un hôpital :

- a) soit il y a eu des malades mais ils ont préféré d'autres hôpitaux pour admission, signifiant l'hôpital peu attractif dans cette localité.
- b) soit parce qu'il n'y a pas eu de malades ou à causes des soins primaires existants, il n'y a pas lieu d'admission ; c'est des zéros qui ne sont pas non liés à l'attractivité de l'hôpital. Donc ceci peut alors nous amener à observer des données avec des excès de zéros.

Dans ce chapitre, nous utilisons pour la modélisation de ces données, des modèles communément appelés Zero-Inflated Models où nous avons pris en compte la possible dépendance spatiale des données au niveau des moyennes locales de la distribution

utilisée. Un des paramètres de la moyenne (celui de la variable distance) sera utilisé pour comparer l'attractivité des hôpitaux et donc les classer.

La première section du chapitre donne un aperçu sur trois Zero-Inflated Models, la seconde section présente un résumé d'application sur les données d'admission et la troisième section, un article révisé soumis à Health Services and Outcomes Research Methodology.

3.1 Les modèles avec des extra-zéros

Dans les données de comptage, il arrive qu'il y ait plus ou d'autres types de zéros que ceux attendus selon la distribution utilisée pour modéliser ces données. Bien sûr il peut exister des cas où il y a moins de zéros que ceux attendus mais c'est rare en pratique et nous ne traitons pas de ce cas dans notre travail du fait des caractéristiques de nos données. Il existe aussi plusieurs distributions pouvant être utilisées pour modéliser ces données mais nous ne ferons cas que de trois modèles : le modèle dit Zero-Inflated Poisson (ZIP), le modèle de Hurdle lié à la distribution de Poisson et le Zero-Inflated Negative Binomial.

3.1.1 Le ZIP et le Hurdle

Quand il s'agit de modéliser des données de comptage, il est souvent naturel d'utiliser la distribution de Poisson. Mais il arrive aussi qu'il y ait plus de zéros dans ces données que réellement de zéros attendus pour une distribution de Poisson. Cet excès de zéros peut être dû à des zéros qui sont structurellement différents du processus de Poisson. Une idée est de mettre une masse p pour une partie des zéros dont on pense ne pas provenir du processus étudié. C'est ainsi que si Y est la variable aléatoire représentant les données, alors le modèle serait

$$Pr(Y = y | p, \lambda) = \begin{cases} p + (1 - p) \exp(-\lambda), & y = 0 \\ (1 - p) \frac{\lambda^y}{y!} \exp(-\lambda), & y > 0. \end{cases} \quad (3.1)$$

Il est possible que p soit négatif, c'est-à-dire moins de zéros que ceux attendus mais comme nous l'avons dit, nous ne considérons que le cas où il y a excès de zéros, donc $0 \leq p < 1$. Ce modèle est appelé Zero-Inflated Poisson (ZIP). Son utilisation remonte aux années 60 avec des auteurs comme Cohen [Cohen, 1963] ou Johnson et Kotz [Johnson and Kotz, 1969]. Plus tard, Lambert [Lambert, 1992] considéra le même modèle avec

$$\log(\lambda) = X\beta \quad \text{et} \quad \log\left(\frac{p}{1-p}\right) = Z\alpha.$$

X , Z sont des matrices de co-variables dont on pense pouvoir améliorer la modélisation et β et α des paramètres à estimer. En ce qui nous concerne, nous avons

considéré λ dépendant du nombre de résidents et de la distance de la localité par rapport à l'hôpital. En plus nous avons supposé ces λ aléatoires et spatialement corrélés.

Une autre approche de modélisation quand il y a des extra-zéros dans les données en cas d'utilisation d'une distribution de Poisson, serait d'utiliser le modèle de Hurdle [Mullahy, 1986]. Il consiste, quant à lui, à donner à tous les zéros une même et unique masse π_0 et à tronquer la distribution pour les valeurs différentes de zéro, c'est-à-dire :

$$Pr(Y = y | \pi_0, \lambda) = \begin{cases} \pi_0, & y = 0 \\ \frac{(1 - \pi_0)\lambda^y \exp(-\lambda)}{\{1 - \exp(-\lambda)\}y!}, & y > 0, \end{cases} \quad (3.2)$$

Y étant la variable aléatoire régissant les données de comptage. En écrivant $\pi_+ = 1 - \pi_0$, des transformations telles que

$$\log(\lambda) = X\beta \quad \text{et} \quad \log\{-\log(1 - \pi_+)\} = Z\alpha$$

peuvent être faites. Même si on peut considérer que $\pi_0 = p + (1 - p)\exp(-\lambda)$, le Hurdle et le ZIP ne sont pas équivalents.

3.1.2 Le modèle ZINB

Une autre distribution qui peut être utilisée pour modéliser des données de comptage, est la binomiale négative. Pour tenir compte des zéros en excès, Greene [Greene, 1994] utilisa le modèle appelé Zero-Inflated Negative Binomial (ZINB) dont voici une variante

$$Pr(Y = y | p, r, \lambda) = \begin{cases} p + (1 - p) \left(1 + \frac{\lambda}{r}\right)^{-r}, & y = 0 \\ (1 - p) \frac{\Gamma(y + r)}{y! \Gamma(r)} \left(\frac{r}{r + \lambda}\right)^r \left(\frac{\lambda}{r + \lambda}\right)^y, & y > 0. \end{cases} \quad (3.3)$$

Les mêmes transformations faites sur p et λ avec le ZIP peuvent être faites sur ce modèle aussi. C'est ce modèle que nous avons utilisé comme alternative au ZIP lors de notre application. Il a la particularité d'être une alternative robuste pour le modèle ZIP. En effet, il est montré que

$$\lim_{r \rightarrow +\infty} \frac{\Gamma(y + r)}{y! \Gamma(r)} \left(\frac{r}{r + \lambda}\right)^r \left(\frac{\lambda}{r + \lambda}\right)^y = \frac{\lambda^y}{y!} \cdot 1 \cdot \frac{1}{\exp(\lambda)},$$

où le terme à droite est l'expression d'une distribution de Poisson. Avec r petit, la variance de ZINB est plus grande que celle de Poisson. En d'autres termes, le ZINB peut bien être utilisé comme alternatif au ZIP.

3.2 Application de ces modèles aux données d'admission de patients

3.2.1 Cadre et précisions

Supposons que nous ayons pour l'étude, une zone (qui peut être un département, une région ...) composée d'un certain nombre de localités (que nous considérons comme des points dans la modélisation confondus à leurs coordonnées géographiques), disons N . Pour question de simplicité, au lieu d'utiliser les noms, nous attribuerons à chacune de ces localités un index i compris entre 1 et N , par exemple selon l'ordre alphabétique de ces noms.

Dans cette zone, existent H hôpitaux dont on veut modéliser les données d'admission. Nous utiliserons des index h compris entre 1 et H au besoin, pour les hôpitaux au lieu de leur nom.

Dans ces hôpitaux sont admis des patients de la zone. Pour admission, le patient peut choisir n'importe quel de ces H hôpitaux sans restriction aucune si ce n'est l'attractivité de l'hôpital telle qu'il la perçoit : coût, qualité des soins, accès facile, commodité ...

Précisons que pour la distance, nous utilisons la distance euclidienne entre coordonnées géographiques de ces localités au lieu de la distance routière que nous aurions pu utiliser sans que cela ne change la méthode.

3.2.2 Modélisation des données d'admission de patients

Nous avons dit que nous disposons pour une pathologie donnée, du nombre d'admissions par localité et par hôpital ; donc des données de comptage avec le nombre (pouvant être petit) de patients par localité admis à un hôpital ; ce nombre est aléatoire car tomber malade est aléatoire. En pareil cas, généralement un modèle de Poisson peut être utilisé (adéquation à montrer). C'est dire que pour une localité i de coordonnées géographiques x_i , le nombre de patients $Y(x_i)$ admis à l'hôpital h suit une loi de Poisson de moyenne λ_h , λ_h étant le nombre moyen d'admission qui plus il est grand, plus l'hôpital est attractif et inversement. Plus précisément, la probabilité d'avoir $Y(x_i) = y(x_i)$ quand λ_h connu, est donnée par

$$P(Y(x_i) = y(x_i) | \lambda_h) = \text{Pois}(Y(x_i) = y(x_i), \lambda_h)$$

où Pois désigne la distribution de Poisson.

Mais les localités n'ont pas le même nombre de résidents ou ne sont pas à la même distance de l'hôpital ; c'est dire que le nombre moyen λ_h peut varier d'une localité à une autre en fonction de la distance (car on a vu que la distance influence le choix des patients) et du nombre de résidents (car le nombre de malades peut varier en fonction du nombre des résidents). Ce qui nous amène à considérer λ_h comme fonction de x_i et de R_i , nombre de résidents de la localité i de telle sorte que

$$\lambda_h(x_i) = C(R_i)\lambda_{0,h}(\|x_i - x_h\|).$$

$C(R_i)$ est une fonction positive et croissante du nombre de résidents R_i telle que $R_i K$, (K désignant la prévalence de la pathologie dans la zone d'étude) désignant intuitivement le nombre de malades de la localité.

Aussi il arrive qu'il y ait des localités avec $Y(x_i) = 0$, non pas parce que les patients n'ont pas choisi cet hôpital mais parce qu'il n'y a pas de patients ou n'ont pas besoin d'admission ; ce qui ne relève pas de l'attractivité de l'hôpital mais plus de l'épidémiologie de la pathologie étudiée. Cette situation conduit à un excès de zéros dans les données. Pour ce faire, une masse $p(x_i)$ (comprise entre 0 et 1) serait affectée pour prendre en compte ce genre de zéros. Mais dans notre étude, nous avons supposé $p(x_i)$ moyen et égal pour toutes les localités pour la raison que nous avons supposé la prévalence constante et il n'y a aucune raison particulière qui puisse faire croire à une grande variabilité des $p(x_i)$ entre les localités, surtout pour un département comme l'Hérault. Ainsi, nous proposons comme modélisation des données d'admission d'un hôpital h :

$$Y(x_i) \mid p, \lambda_h(x_i) \sim \begin{cases} p + \{1 - p\} \exp\{-\lambda_h(x_i)\}, & Y(x_i) = 0 \\ \{1 - p\} \frac{\lambda_h(x_i)^{Y(x_i)}}{Y(x_i)!} \exp\{-\lambda_h(x_i)\}, & Y(x_i) > 0. \end{cases} \quad (3.4)$$

Par ailleurs, la loi binomiale négative constituant une alternative robuste à une loi de Poisson, une modélisation alternative à la précédente serait :

$$Y(x_i) \mid p, r, \lambda_h(x_i) \sim \begin{cases} p + \{1 - p\} \left\{1 + \frac{\lambda_h(x_i)}{r}\right\}^{-r}, & Y(x_i) = 0 \\ \{1 - p\} \frac{\Gamma\{Y(x_i)+r\}}{Y(x_i)! \Gamma\{r\}} \left\{\frac{r}{r+\lambda_h(x_i)_h}\right\}^r \left\{\frac{\lambda_h(x_i)}{r+\lambda_h(x_i)}\right\}^{Y(x_i)}, & Y(x_i) > 0. \end{cases} \quad (3.5)$$

Il faut noter que nous avons considéré, dans l'article, le même r pour tous nos 4 hôpitaux en compétition du fait d'une étude préliminaire (non présentée ici) ayant montré que r est sensiblement le même pour ces 4 hôpitaux.

3.2.3 Amélioration de la modélisation des données par prise en compte de la dépendance spatiale

Nous avons supposé les λ_h variant d'une localité à une autre. Aussi les populations voisines partagent informations et habitudes. De ce fait, les raisons qui ont conduit les patients d'une localité au choix d'un hôpital donné peuvent être semblables à celles des patients d'une localité voisine ; il existe donc une corrélation entre les nombres d'admissions de ces localités. L'ajout d'une couche à (3.4) ou (3.5) prenant en compte cette dépendance spatiale peut améliorer la modélisation des données d'admission. En pareil cas, le processus gaussien est généralement utilisé. Comme les $\lambda_h(x_i)$ doivent être positifs et peuvent être le résultat de la conjonction de plusieurs

facteurs, nous avons utilisé le processus log-normal qui prend mieux en compte cette spécificité. Ainsi, si par exemple, $\mathbf{x} = (x_1, \dots, x_N)$, $\lambda_h(\mathbf{x}) = (\lambda_h(x_1), \dots, \lambda_h(x_n))'$ et $R = (R_1, \dots, R_N)$, alors

$$\lambda_h(\mathbf{x}) \mid \mu_h(\mathbf{x}), \gamma_h(\mathbf{x}), C(R) \sim (2\pi)^{-n/2} |\gamma_h(\mathbf{x})|^{-1/2} \times \prod_{i=1}^n \lambda_h(x_i)^{-1} \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\}, \quad (3.6)$$

où $\mu_h(\mathbf{x})$ est la moyenne du processus et $\gamma_h(\mathbf{x}) = (\gamma_{ij,h})_{i,j}$, sa matrice de covariance.

Différentes formes fonctionnelles (linéaire, polynomiale...) peuvent être supposées pour $\mu_h(x_i)$ et $C(R_i)$ et différentes fonctions de covariance pour $(\gamma_{ij,h})_{i,j}$. Dans notre étude, nous avons considéré que

$$\gamma_{ij,h} = \tau_h \exp \left(-\frac{\|x_i - x_j\|}{\omega_h} \right),$$

(famille exponentielle isotrope, choix plus guidé par la simplicité de la fonction) et $C(R_i) = R_i K$, représentant intuitivement le nombre de malades de la localité i .

Pour définir $\mu_h(x_i)$, nous avons utilisé le fait que

$$\log\{\lambda_h(x_i)\} = \log\{C(R_i)\} + \log\{\lambda_{0,h}(x_i)\}$$

impliquant que

$$\log\{\lambda_{0,h}(\mathbf{x})\} \sim N\{\mu'_h(\mathbf{x}), \gamma_h(\mathbf{x})\}$$

tel que $\mu_h(\mathbf{x}_i) = \mu'_h(\mathbf{x}_i) + \log\{C(R_i)\}$. Nous avons supposé que

$$\mu'_h(x_i) = \beta'_{0,h} + \beta_{1,h} \|x_i - x_h\|,$$

fonction linéaire (qui même si elle est restrictive, peut être considérée comme première approximation d'une forme fonctionnelle régulière) de la distance euclidienne $\|x_i - x_h\|$ séparant l'hôpital de la localité du patient. On obtient alors

$$\mu_h(x_i) = \beta'_{0,h} + \log(R_i) + \log(K) + \beta_{1,h} \|x_i - x_h\|,$$

c'est-à-dire

$$\mu_h(x_i) = \beta_{0,h} + \log(R_i) + \beta_{1,h} \|x_i - x_h\|,$$

forme linéaire de la distance où $\beta_{0,h} = \beta'_{0,h} + \log(K)$, nous permettant d'éviter d'estimer $\beta'_{0,h}$ et K mais seulement $\beta_{0,h}$.

$\beta_{1,h}$ est le paramètre qui contrôle comment varie $\mu_h(x_i)$, donc $\lambda_h(x_i)$ au fur et à mesure qu'on s'éloigne de l'hôpital : plus il est grand, plus la moyenne $\lambda_h(x_i)$ serait grande, c'est-à-dire l'hôpital plus attractif. Il sera utilisé pour la comparaison des hôpitaux.

Une dépendance spatiale aurait pu être utilisée pour les $p(x_i)$ et K mais dans notre article nous les avons considérés constants, comme moyennes dans la zone d'étude.

Pour l'estimation des paramètres, nous avons choisi un cadre bayésien et les méthodes MCMC.

3.3 Article

Health Services and Outcomes Research Methodology

Comparing hospitals spatial attractiveness by using zero-inflated spatial models

--Manuscript Draft--

Manuscript Number:	HSOR-D-16-00038R2
Article Type:	Manuscript
Keywords:	Hospital Attractiveness; Spatial Bayesian Hierarchical Model; Zero-Inflated Data.
Corresponding Author:	Issa SALEY Universite de Montpellier Montpellier, FRANCE
First Author:	Issa SALEY
Order of Authors:	Issa SALEY Nicolas Molinari Mathieu Ribatet
Funding Information:	
Response to Reviewers:	<p>T*****he associate editor has reviewed your revisions, recognizes the strengths of the paper, but cannot approve the paper in its current form. While the math in the paper is informative and concise the writing fails to communicate the main considerations and justify the fundamental modelling decisions. It took me a considerable amount of time to justify the model to myself, as the writing failed to be compelling.*****</p> <p>We acknowledge that the justifications were not clearly given, In this version we provide justifications such as :</p> <p>Our first model is based on a Poisson distribution to model the number of admissions in a given hospital at a given location. At first sight the use of the Poisson distribution seems particularly relevant as the probability that a single person at risk gets ill and decides to go to a given hospital is rather small and hence justifies considering, at least theoretically, the so-called "law of small numbers"..... and</p> <p>A widely used alternative to the zero inflated Poisson model is the zero inflated negative binomial model. As a mixture of Poisson and Gamma distributions, there are some situations where the negative binomial distribution is better suited than the Poisson distribution</p> <p>*****The paper also fails to emphasize the intercepts themselves, that is the attraction of the hospital when one lives 0 distance from the hospital. It could be that a hospital is extremely popular locally but unknown farther away.*****</p> <p>To take this comment into consideration , our response is :</p> <p>Hospital attractiveness can be achieved by inspecting the posterior distribution of β_{0h} and β_{1h}. Of primary importance is β_{1h} as it controls how varies the log-mean of the process as one goes away to the hospital while β_{0h} characterizes attractiveness at "home city". For example, from</p>

Table 2 one can see that Clinique le Millénaire is the less attractive at home city, due probably to the presence of CHU Montpellier in the same city. CH Béziers seems to be more attractive at home city; it is practically the only well-known center in the city of Béziers.

*****The paper also fails to address capacity of the hospital*****|

Since patient may take into consideration accomodation (such availability of beds) before choosing a hospital, we consider capacity of the hospital as part of attractiveness of the hospital. Thus ,in introduction we wrote : Indeed in regions where several hospitals can be chosen by the patients, it seems reasonable to think that the selected one meets the patient's needs better than the other hospitals, e.g., price, quality of health care, ease of access, accommodation.

*****There is too many tables. Please remove the tables on R.hat statistics, and only refer to them in the text*****

The tables on R.hat statistics are removed.

*****The importance of Figure 3 is not clear. I would revise it to be histograms or density estimates of distance from hospital of admitted patients for each of the 4 hospitals.*****

The figure is removed since its importance is not clear.

*****It is also recommended that the authors get an English Language Editing Service to help improve scientific writing*****

English is checked by a writer fluent in English.

[Click here to view linked References](#)

Comparing the spatial attractiveness of hospitals using zero-inflated spatial models

I. Saley

IMAG, UMR CNRS 5149, Université de Montpellier, 4 place Eugène Bataillon,
34095 Montpellier Cedex 5, France

N. Molinari

Service DIM, CHRU de Montpellier, 39 avenue Charles Flahaut,
34 295 Montpellier Cedex 5, France

M. Ribatet

IMAG, UMR CNRS 5149, Université de Montpellier, 4 place Eugène Bataillon,
34095 Montpellier Cedex 5, France

Abstract

Policy makers increasingly rely on hospital competition to incentivize patients to choose high-value care. Amongst all possible drivers, the travel distance without any doubt is one of the most important. In this paper we propose the use of a spatial Bayesian hierarchical model to assess the impact of distance on the number of patient admissions in hospitals, and thereby, compare hospital attractiveness. To this aim a MCMC sampler has been designed. We apply our methodology to patient admissions for asthma in four hospitals located in the Hérault department of France. Results indicate that the most attractive hospital is the CHU Montpellier.

Keywords: Hospital Attractiveness; Spatial Bayesian Hierarchical Model; Zero-Inflated Data

1 Introduction

Competition between hospitals for the provision of health care has been increasingly advocated for its potential to improve efficiency as well as the quality of care (Propper and Leckie, 2011). As a consequence of this increasing competition, it is hoped that hospitals decrease production costs and improve the quality of health care delivered so as to attract new patients. The success of this strategy depends on the patients' propensity to switch hospitals in order to choose high-value care. Understanding how patients choose hospitals is, therefore, of major importance for health care providers and policy makers. Hospital attractiveness factors include price, quality, distance (or travel time), waiting time, provider network and others. Early studies identified distance or travel time as the major factor negatively affecting hospital choice, including in metropolitan areas where there are many hospitals within short distances (Victoor et al., 2012), even though the sensitivity to distance varies with patient characteristics (age, ethnicity, income and religion) and the type of admission.

As in many European countries, French hospitals are financed through Diagnosis Related Groups based on a prospective payment system. Regarding acute care, this system was fully implemented in 2005 for private-for-profit hospital budgets and in 2008 for public hospital budgets. The reform was intended to improve efficiency and fairness in financing and also to increase competition between and within the public and private sectors (Chevreul et al., 2010).

Information on hospital performance and quality in France hasn't been published yet. Hospital selection is made by the patients themselves, following advice from their general practitioners.

1
2
3
4
5
6
7
8 Since general practitioners do not face any financial incentives to refer their patients to a given
9 hospital, we can assume that they take into account patient preferences, among other factors.
10 Thus, we can hypothesize that reputation (Jung et al., 2011), as perceived by the patients and the
11 general practitioners, partially reflects the quality of care. In these circumstances, travel distance
12 certainly plays a key role in patient decisions (Varkevisser et al., 2010).

13 In this context, spatial models describing the number of patient admissions may be useful
14 for hospital classification (from most to least attractive). **Indeed in regions where several
15 hospitals can be chosen by the patients, it seems reasonable to think that the selected
16 one meets the patient’s needs better than the other hospitals, e.g., price, quality of
17 health care, ease of access, accommodation.** . . In particular, when patients tend to visit a
18 hospital regardless of the distance traveled, it seems reasonable to assume that this hospital can
19 be tagged as “very attractive”. From a policy maker standpoint, such conclusions could offer a
20 framework for regulation.

21 In this paper, we introduce a spatial Bayesian hierarchical model to assess the impact of
22 distance on the number of patient admissions. In the first layer of the model, i.e., *the data layer*,
23 the number of patients admitted to a specific hospital at a given location is modelled by a Poisson
24 distribution whose mean drives how attractive is this hospital. Note however that it is possible
25 that no admission occurs. Typically two different cases may cause such behaviour:

- 26 • all patients have decided to go to other hospitals—and hence indicates poor hospital attrac-
27 tiveness;
- 28 • all potential subjects are “healthy” or have already been healed by primary health facilities, so
29 that in both cases, there is no need for an “admission process”. In such situation, evaluation
30 of the hospital attractiveness is impossible.

31 To cope with this problem, a common practice is to consider zero inflated models, e.g., Zero
32 Inflated Poisson (ZIP), where a mass p is considered to capture these extra-zeros (Cohen, 1963;
33 Johnson and Kotz, 1969). (Lambert, 1992) used a ZIP model where both the p probability and
34 the mean λ of the Poisson distribution can depend on covariates. In this paper we propose to
35 analyse in addition to covariate dependent ZIP model, Zero Inflated Negative Binomial (ZINB)
36 models (Greene, 1994).

37 Another key feature when modelling patients’ admission is that the data are intrinsically spatial
38 as a given disease is likely to impact a town as well as its neighbouring cities. (Agarwal et al.,
39 2002) used a spatial zero-inflated Poisson model for areal data where the Poisson mean is modelled
40 by a standard log-linear model. They modelled the spatial dependence of the data within the
41 random effects that they imposed in the standard log-linear model by a Gaussian Markov field.
42 In this paper, we suppose that the mean of the Poisson distribution, $\lambda(x)$, depends on patient
43 residence location x . In the second layer, i.e., *the process layer*, we model $\lambda(x)$ as random and
44 assume a multivariate log-normal distribution whose mean vector depends on relevant covariates,
45 e.g., distance between the patient location x and the hospital, and covariance matrix is derived
46 from an exponential correlation function family. Finally as we are working within a Bayesian
47 framework prior distributions are assumed to each parameters of the model. As often with Bayesian
48 hierarchical models, inference is based on the posterior distribution which can be obtained using
49 Markov Chain Monte Carlo (MCMC) methods.

50 Section 2 describes the data set analysed in this paper. Section 3 introduces the statistical
51 models under study while section 4 describes the implementation details for our MCMC sampler.
52 Section 5 presents results obtained with an illustrative data set. The paper ends with a brief
53 discussion and supplementary material about the MCMC sampler can be found in Section 4.

54 2 The Asthma data set

55 The principal data source used in this paper is the *Programme de Médicalisation des Systèmes*
56 *d’Information* (PMSI). The PMSI is an administrative data set recording all patient admissions in
57
58
59
60
61
62
63
64
65

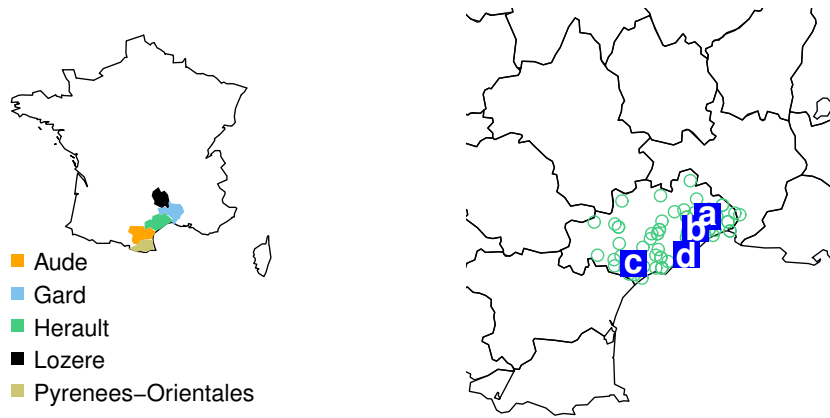


Figure 1: Mapping of the illustration data. The left map shows where the Languedoc-Roussillon Region is located in France and subsequently the Hérault department. The picture on the right displays hospital locations (a: CHU Montpellier, b: Clinique le Millénaire, c: CH Béziers and d: CH Bassin de Thau) and localities (green points) where at least one resident admission is reported in these hospitals in the study.

Table 1: Number of admission per hospital for the Asthma data set.

Hospital	number of asthma patient stays
CH Bassin de Thau	71
CH Béziers	79
CHU Montpellier	127
Clinique Le Millénaire	23

French private and public hospitals covering all social health insurance programs. In this data set, limited information about the patient and the hospital stay, such as the disease, place of residence, location of the hospital, type of hospital, and travelling distance, are recorded. Therefore, for every region or department in France, we can obtain the number of patients per locality admitted in a given hospital for a given disease.

In this study we focus on the Languedoc–Roussillon region of France. The left panel of Figure 1 highlights this region which is composed of five departments: Aude, Gard, Hérault, Lozère and Pyrénées–Orientales. In 2009, 407 hospital patient stays were recorded for asthma in Hérault hospitals (see PMSI database). The department consists of 343 localities called “communes” (municipalities) spread over 6224 km² with a population of approximately 1.1 million, including both rural and metropolitan areas. The Hérault department is known to have a high competitive health care market in France.

In this paper we focus on the main 4 hospitals of Hérault, i.e., with the largest number of asthma patient stays. Among these four medical institutions, only one is private-for-profit while the remaining ones are not-for-profit. The right panel of Figure 1 plots the locations of these four hospitals as well as the spatial locations of localities where at least one resident admission in these hospitals is reported. Table 1 displays the distribution of the number of admission per hospital.

3 Models

As a reminder, the aim of this paper is to provide a method for measuring the effect of distance on the spatial number of patient admissions for a given hospital by using a spatial Bayesian hierarchical model, and to subsequently compare their attractiveness. We suppose that a patient can choose any of the hospitals in competition without any restriction than the way he or she

perceives the attractiveness of the hospital.

For the purpose of this study, we suppose that for each localities and each hospital, we have access to the number of admissions. The number of residents per locality as well as the pairwise distance matrix between localities are available. Note however that for the distance between localities, the Euclidean distance was used instead of road distance. Indeed as our application involves a large number of localities, obtaining pairwise road distances between localities is difficult.

3.1 Data layer

Our first model is based on a Poisson distribution to model the number of admissions in a given hospital at a given location. At first sight the use of the Poisson distribution seems particularly relevant as the probability that a single person at risk gets ill and decides to go to a given hospital is rather small and hence justifies considering, at least theoretically, the so-called “law of small numbers”. Let $Y_h(x_i)$ denotes the number of admissions registered for hospital $h \in \{1, \dots, H\}$ at the i -th locality, i.e., whose geographic coordinates are x_i . We assume that $Y_h(x_i)$ follows a Poisson distribution with mean λ_h where λ_h expresses how attractive the hospital is for patients. Since it is sensible to assume that the function $x \mapsto \lambda_h(x)$ varies smoothly in space, we assume that

$$\lambda_h : x \mapsto C(R_i)\lambda_{0,h}(\|x_i - x_h\|),$$

for some positive and non-decreasing function C and where R_i denotes the number of resident for the i -th locality. One possibility for C is to take $C(R_i) = KR_i$, $K > 0$, where K is can be thought as the overall probability to be ill, i.e., the so called prevalence ratio, so that $C(R_i)$ would represent the expected number of resident affected by asthma independently of the geographic location. The function $d \mapsto \log \lambda_{0,h}(d)$ is typically taken to be a polynomial.

As stated in the introduction, the number of admissions, $Y_h(x_i)$, can be equal to zero in two different cases: a lack of attractiveness of hospital h (all patients have chosen to go to other hospitals); nobody is affected by asthma in the i -th locality or have been healed by primary health facilities so that there is no need for an “admission process”. This feature induces an extra probability mass at 0 that needs to be taken into account in our modelling strategy. Consequently, it makes sense to use a mass $p(x_i)$ depending on locality i for zeros not related to attractiveness so that the first layer of our model becomes

$$Y_h(x_i) \mid p(x_i), \lambda_h(x_i) \sim \begin{cases} p(x_i) + \{1 - p(x_i)\} \exp\{-\lambda_h(x_i)\}, & Y_h(x_i) = 0 \\ \{1 - p(x_i)\} \frac{\lambda_h(x_i)^{Y_h(x_i)}}{Y_h(x_i)!} \exp\{-\lambda_h(x_i)\}, & Y_h(x_i) > 0, \end{cases} \quad (1)$$

where $k \mapsto k!$ corresponds to the factorial function.

A widely used alternative to the zero inflated Poisson model is the zero inflated negative binomial model. As a mixture of Poisson and Gamma distributions, there are some situations where the negative binomial distribution is better suited than the Poisson distribution. Keeping the same notations as in (1), the data layer for this model is

$$Y_h(x_i) \mid p(x_i), r, \lambda_h(x_i) \sim \begin{cases} p(x_i) + \{1 - p(x_i)\} \left\{1 + \frac{\lambda_h(x_i)}{r}\right\}^{-r}, & Y_h(x_i) = 0 \\ \{1 - p(x_i)\} \frac{\Gamma\{Y_h(x_i)+r\}}{Y_h(x_i)!\Gamma(r)} \left\{1 + \frac{\lambda_h(x_i)}{r}\right\}^{-r} \left\{\frac{\lambda_h(x_i)}{r+\lambda_h(x_i)}\right\}^{Y_h(x_i)}, & Y_h(x_i) > 0, \end{cases} \quad (2)$$

where $\Gamma(\cdot)$ is the Gamma function. One recover the zero inflated Poisson model as $r \rightarrow \infty$.

3.2 Process layer

It is sensible to assume that residents from neighbouring localities will have and share the same habits and decisions. For instance, persons can take into consideration what their neighbours

1
2
3
4
5
6
7
8 opinions/experience before choosing a hospital. It is therefore reasonable to assume that the
9 random vector $\lambda_h(\mathbf{x}) = \{\lambda_h(x_1), \dots, \lambda_h(x_n)\}^\top$ exhibits some kind dependence. Mathematically
10 speaking we assume that $\lambda_h(\mathbf{x})$ follows a multivariate log-normal distribution so that one can
11 benefit of the so-called *borrowing strength* of hierarchical models to characterize $\lambda_h(x_i)$ for some
12 $i \in \{1, \dots, n\}$. We then have

$$13 \lambda_h(\mathbf{x}) \mid \mu_h(\mathbf{x}), \gamma_h(\mathbf{x}), C(R) \sim (2\pi)^{-n/2} |\gamma_h(\mathbf{x})|^{-1/2} \times \\ 14 \prod_{i=1}^n \lambda_h(x_i)^{-1} \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]^\top \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\}, \quad (3)$$

15 where $\mu_h(\mathbf{x})$ is the mean vector of the log-normal distribution and $\gamma_h(\mathbf{x}) = \{\gamma_h(x_i - x_j)\}_{i,j=1,\dots,n}$
16 its covariance matrix.

17 Since $\log \lambda_h(x_i) = \log C(R_i) + \log \lambda_{0,h}(x_i)$, we have conditionally on $C(R_i)$, $\log \lambda_{0,h}(\mathbf{x}) \sim$
18 $N\{\mu'(\mathbf{x}), \gamma_h(\mathbf{x})\}$ with $\mu(\mathbf{x}_i) = \mu'(\mathbf{x}_i) + \log C(R_i)$.

19 Throughout this study, we assume that

$$20 \mu'_h(x_i) = \beta'_{0,h} + \beta_{1,h} \|x_i - x_h\|, \quad i = 1, \dots, n, \quad h = 1, \dots, H,$$

21 where x_h denotes the geographic coordinates of hospital $h \in \{1, \dots, H\}$. This implies that, for
22 any $i \in \{1, \dots, n\}$,

$$23 \mu_h(x_i) = \beta'_{0,h} + \log R_i + \log K + \beta_{1,h} \|x_i - x_h\| \\ 24 = \beta_{0,h} + \log R_i + \beta_{1,h} \|x_i - x_h\|,$$

25 where $\beta_{0,h} = \beta'_{0,h} + \log K$ so that there is no need to impose any distribution on the unknown
26 prevalence ratio K .

27 Among all parameter, $\beta_{1,h}$ is of primary importance for our study as it controls how the log-
28 mean of the Poisson/Negative binomial distribution varies with the distance from the hospital and
29 therefore summarize to some extent how attractive is this hospital. As a consequence it is expected
30 that the larger the $\beta_{1,h}$, the more attractive is hospital h . We can rank hospital attractiveness by
31 ranking their respective summary of the posterior distribution of $\beta_{1,h}$, e.g., posterior mean.

32 To avoid over pasteurization, it is often more convenient to add parametric structure for the
33 covariance function. In this study we assume an isotropic exponential covariance function family
34 for $\gamma_h(\mathbf{x})$, i.e.,

$$35 \gamma_h(x_i - x_j) = \tau_h \exp \left(-\frac{\|x_i - x_j\|}{\omega_h} \right),$$

36 where $\tau_h > 0$ is the variance and $\omega_h > 0$ the scale parameter. It is well known that consid-
37 ering covariance function families with a shape parameter such as the powered exponential or
38 Whittle–Matèrn induces some non-identifiability issues (Zhang, 2004; Sang and Gelfand, 2010;
39 Davison et al., 2012). The exponential covariance family appear to be a good trade off between
40 identifiability problems and flexibility.

41 Although a spatial correlation could be assumed on $p(x_i)$ and to avoid over parametrized
42 models, we treat p as spatially constant and independent of the hospital so that we will write
43 p instead of $p(x_i)$. A preliminary study, not presented here, showed that the r parameter was
44 mainly the same for all 4 hospitals. Consequently throughout this paper we will consider a zero
45 inflated negative binomial model where the r parameter is constant across hospitals. To have
46 a complete picture of the proposed models, Figure 2 gives the direct acyclic graph for the zero
47 inflated Poisson/negative binomial models.

48 3.3 Parameters prior distributions

49 Since we are working in a Bayesian framework, prior distributions have to be assumed on each
50 parameters of the model. In this study we assign independent priors for each parameters, i.e.,

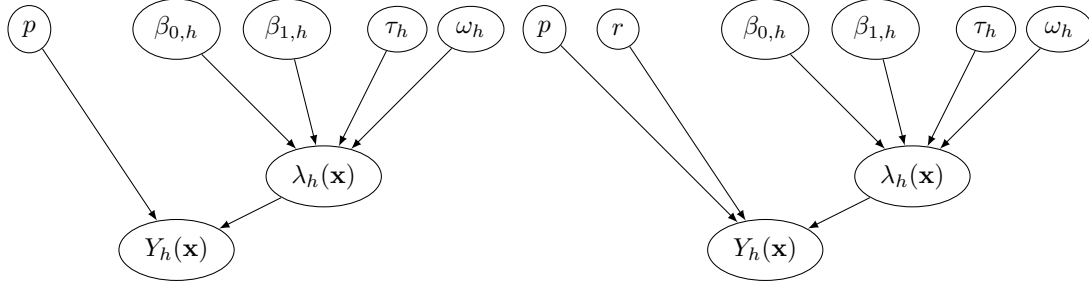


Figure 2: Direct acyclic graph of the Bayesian hierarchical models. Left: Zero inflated Poisson model. Right: Zero inflated negative binomial model.

p , r , $\beta_{0,h}$, $\beta_{1,h}$, τ_h and ω_h . For numerical reasons we use a reparametrization for p by letting $p = 1/(1 + e^{-\theta})$ and thus assign a Gaussian prior distribution on θ instead of p . Whenever possible we use conjugate priors to speed up our MCMC sampler. More precisely a non informative Gaussian prior distribution is assumed for $\beta_{0,h}$ and $\beta_{1,h}$. For the zero inflated negative binomial model a uniform distribution on the interval $(0, 100)$ was assumed for r . More care is need when defining prior distributions for τ_h and ω_h . Indeed (Berger et al., 2001) and (Banerjee et al., 2004) showed that improper priors for these parameters, can, in general, result in improper posterior distributions. Following (Banerjee et al., 2004) who suggested using informative prior distributions, we assumed an Inverse-Gamma distribution with shape and scale parameters equal to 2 for both τ_h and ω_h . As this specific choice for the prior distribution is somehow arbitrary, we conducted a sensitivity analysis about the impact of such prior distribution in Section 5.2.

4 Full conditional distribution

To ease the notation let $\lambda(\mathbf{x}) = \{\lambda_1(\mathbf{x}), \dots, \lambda_H(\mathbf{x})\}$, $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,H})$, $\beta_1 = (\beta_{1,1}, \dots, \beta_{1,H})$, $\tau = (\tau_1, \dots, \tau_H)$ and $\omega = (\omega_1, \dots, \omega_H)$. Similarly we write $\lambda_{-h}(\mathbf{x})$, $\beta_{0,-h}$, $\beta_{1,-h}$, τ_{-h} and ω_{-h} to denote the corresponding vector without the h -th element. As often with hierarchical models, the likelihood involves multiple integrals so that likelihood-based inference is not possible. To bypass this hurdle, a widely used strategy is to have resort to Monte Carlo Markov Chain algorithms whose aim is to generate a Markov chain whose stationary distribution is the posterior distribution.

For our purposes, a particularly well suited MCMC algorithm is the Gibbs sampler (Gelfand and Smith, 1990). This sampler is based on the full conditional distributions which turn out to be for the zero inflated Poisson model

$$\begin{aligned} \pi\{\theta \mid y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} &\propto \pi\{y(\mathbf{x}) \mid \theta, \lambda(\mathbf{x})\}\pi(\theta) \\ \pi\{\lambda_h(\mathbf{x}) \mid y(\mathbf{x}), \lambda_{-h}(\mathbf{x}), \theta, \beta_0, \beta_1, \tau, \omega\} &\propto \pi\{y(\mathbf{x}) \mid \theta, \lambda(\mathbf{x})\}\pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \\ \pi\{\beta_{0,h} \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_{0,-h}, \beta_1, \tau, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\beta_{0,h}) \\ \pi\{\beta_{1,h} \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_{1,-h}, \tau, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\beta_{1,h}) \\ \pi\{\tau_h \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau_{-h}, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\tau_h) \\ \pi\{\omega_h \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega_{-h}\} &\propto \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\omega_h). \end{aligned}$$

Similar expressions are readily obtained for the zero inflated negative binomial model.

Note that since many of the above full conditional distributions are not related to well identified distribution such as Gaussian or Gamma distributions, Metropolis-Hasting updating scheme might be used in various places (Gelman et al., 2004). When this is the case a random walk, eventually of the log-scale for non-negative parameters, with Gaussian innovation was used. This Metropolis-Hastings within Gibbs algorithm was written in R (R Core Team, 2017) and could be provided upon request to the authors.

Table 2: Posterior means for the ZIP/ZINB parameters and associated Standard Errors (SE) obtained from 10 MCMC chains for the asthma data set. The Deviance Information Criterion (DIC) for each model is also reported.

ZIP Model: $p = 0.58$ (0.02), DIC = 653 (6)					
	$\beta_{0,h}$	$\beta_{1,h}$	τ_h	ω_h	τ_h/ω_h
CH Bassin de Thau	-8.18 (0.26)	-0.058 (0.009)	1.03 (0.32)	2.5 (2.0)	0.91 (0.14)
CH Béziers	-6.99 (0.15)	-0.052 (0.004)	1.09 (0.29)	2.5 (2.5)	0.93 (0.15)
CHU Montpellier	-8.21 (0.14)	-0.015 (0.002)	0.92 (0.15)	2.4 (1.4)	0.95 (0.16)
Clinique Le Millénaire	-9.56 (0.11)	-0.019 (0.002)	0.85 (0.20)	1.4 (0.4)	0.92 (0.18)
ZINB Model: $p = 0.59$ (0.02), $r = 47.2$ (2.7), DIC = 669 (8)					
Hospital	$\beta_{0,h}$	$\beta_{1,h}$	τ_h	ω_h	τ_h/ω_h
CH Bassin de Thau	-8.04 (0.14)	-0.055 (0.010)	1.04 (0.29)	2.76 (1.23)	0.804 (0.17)
CH Béziers	-7.05 (0.10)	-0.055 (0.004)	1.04 (0.23)	1.98 (0.63)	0.95 (0.19)
CHU Montpellier	-8.34 (0.05)	-0.012 (0.002)	0.94 (0.09)	2.16 (1.20)	0.89 (0.17)
Clinique Le Millénaire	-9.76 (0.11)	-0.018 (0.002)	0.77 (0.08)	1.61 (0.90)	0.82 (0.20)

5 Application to the Asthma data set

In this section, we apply the methodology introduced in Section 3 and 4 to the Asthma data of Section 2. Recall that our analysis focuses on 4 hospitals: CH Bassin de Thau in the city of Sète, CH Béziers in the city of Béziers and both CHU Montpellier and Clinique le Millénaire in the city of Montpellier. The localities used for estimation are all the 343 localities within the Hérault department of France.

5.1 Results

5.1.1 MCMC chains and estimates

For convergence control and parameter estimation, 10 Markov chains of length 50,000 were generated. For each chain, the first 30,000 iterations were treated as burn-in-period and a thinning lag of 10 iterations was used to mitigate serial dependence in the simulated Markov chains. At the end of this post-processing of our MCMC sampler, a chain of length 2,000 that could be reasonably considered as sampled from the posterior distribution and showing very weak serial dependence was used. To double check if convergence to the stationary distribution was reached, the \hat{R} statistics (Aho, 2015) was used. Results indicates that convergence to the stationary distribution, i.e., the posterior distribution, can be safely assumed.

Table 2 reports the estimated posterior means and its standard error obtained from these 10 Markov chains.

5.1.2 Model selection

One first question of interest is to determine if one should prefer the zero inflated negative binomial model over its Poisson counterpart. To choose the model that better fits our data from among the ZIP and ZINB models, we used the Deviance Information Criterion (DIC), (Spiegelhalter et al., 2002). The smaller the DIC, the better the model fits the data. These DIC values are reported in Table 2 as well as their respective standard errors. Although the difference is small, we can see that the DIC value tends to be in favour of the ZIP model. This conclusion could be expected since the posterior mean for r in the ZINB model is large and it is well known that the negative binomial distribution converges to the Poisson distribution as $r \rightarrow \infty$. Consequently and in accordance with Ockam's razor, the DIC tends to prefer the simpler model, i.e., Poisson.

To confirm this conclusion we also analyse prediction performance for each model based on the

Table 3: Mean squared error (MSE) and associated standard errors (in bracket) for the ZIP and ZINB models on the Asthma data set.

	ZIP	ZINB
CH Bassin de Thau	1.71 (1.09)	2.87 (1.31)
CH Béziers	0.85 (0.20)	1.02 (0.28)
CHU Montpellier	0.39 (0.10)	0.44 (0.08)
Clinique Le Millénaire	0.10 (0.02)	0.13 (0.04)

Mean Squared Error criterion

$$MSE_h = \frac{1}{343} \sum_{i=1}^{343} \{y_h(x_i) - \hat{y}_h(x_i)\}^2, \quad h = 1, \dots, H.$$

where $y_h(x_i)$ is the observed value and $\hat{y}_h(x_i)$ the computed posterior mean at locality i . Table 3 shows the computed MSE for both models. Although both models show good performance, the MSE criteria tends to be in favour our the ZIP model and is in agreement with our previous conclusions. Based on these results we will therefore consider only the ZIP model for the rest of this study.

5.1.3 Comparing hospital attractiveness

As stated in Section 3, hospital attractiveness can be achieved by inspecting the posterior distribution of $\beta_{0,h}$ and $\beta_{1,h}$. Of primary importance is $\beta_{1,h}$ as it controls how varies the log-mean of the process as one goes away to the hospital while $\beta_{0,h}$ characterizes attractiveness at “home city”, i.e., when $\|x - x_h\| = 0$. For example, from Table 2 one can see that Clinique le Millénaire is the less attractive at home city, due probably to the presence of CHU Montpellier in the same city. CH Béziers seems to be more attractive at home city; it is practically the only well-known center in the city of Béziers.

Table 2 shows that the posterior mean of $\beta_{1,h}$ are negative for all hospitals. This indicates that the mean of the Poisson distribution, or equivalently hospital attractiveness, decreases as one moves away from the hospital. Such behaviour could be expected since, as identified by prior studies (Victoor et al., 2012), distance is a factor that negatively affects hospital choice. Although the Clinique Le Millénaire is slightly less attractive, the CHU Montpellier appears to be the most attractive hospital. Such findings could be expected since the CHU Montpellier is a well-known, large public health center and has the largest accommodation capacity in the region. It is also known for delivering high quality care: the faculty of medicine from which it depends, is one of the oldest in Europe (founded in the 13th century). The city of Montpellier also has easy access: highway A9, railways (TGV and TER) and many other roads for access. The less attractive hospital is the CH Bassin de Thau for which more than half of admissions are from the home city (Sète) and only five localities out of 343 do not have zero admissions. This behaviour can be explained by the fact that the CH Bassin de Thau is a small public center and is located between Montpellier and Béziers where the 2 largest public centers in the department are found. Sète also lacks the transportation facilities available in the other 2 cities : no TGV or highway. Despite its 22 admissions, the Clinique le Millénaire seems to draw patients from farther away than CH Bassin de Thau and CH Béziers. The Clinique Le Millénaire is a well-known private center in Montpellier and its neighbourhood. It has a good reputation. The small number of admissions can be explained by a smaller accommodation capacity and it is, like other private centers, more specialized in acute health care such as surgery. It is also in the same city with the biggest well-known center: CHU Montpellier. Globally, the proposed center ranking seems sensible.

Table 4: Posterior means for ZIP parameters and associated Standard Errors over 10 MCMC chains for the asthma data set as the prior distributions for τ_h and ω_h vary.

	Prior: Inverse-Gamma(3,3), $p = 0.574$ (0.021)			
	$\beta_{0,h}$	$\beta_{1,h}$	τ_h	ω_h
CH Bassin de Thau	-8.07 (0.14)	-0.057 (0.003)	0.99 (0.25)	1.68 (0.63)
CH Béziers	-7.04 (0.03)	-0.054 (0.002)	1.21 (0.15)	1.75 (0.41)
CHU Montpellier	-8.42 (0.03)	-0.016 (0.002)	0.95 (0.08)	1.27 (0.22)
Clinique Le Millénaire	-9.73 (0.11)	-0.016 (0.003)	0.94 (0.08)	1.25 (0.17)

	Prior: Inverse-Gamma(1,1), $p = 0.576$ (0.013)			
	$\beta_{0,h}$	$\beta_{1,h}$	τ_h	ω_h
CH Bassin de Thau	-8.13 (0.17)	-0.056 (0.004)	1.00 (0.35)	37.24 (4.69)
CH Béziers	-7.11 (0.07)	-0.053 (0.003)	1.11 (0.22)	6.92 (4.26)
CHU Montpellier	-8.24 (0.04)	-0.015 (0.002)	0.96 (0.21)	4.38 (3.68)
Clinique Le Millénaire	-9.82 (0.05)	-0.016 (0.005)	0.84 (0.26)	2.61 (2.18)

Table 5: Table of posterior means for the ratio τ_h/ω_h and associated Standard Errors (SE) over 10 MCMC chains as the prior distributions vary.

	Inverse-Gamma (1,1)	Inverse-Gamma(2,2)	Inverse-Gamma(3,3)
CH Bassin de Thau	0.55 (0.35)	0.91 (0.14)	0.86 (0.13)
CH Béziers	0.71 (0.32)	0.93 (0.15)	1.06 (0.07)
CHU Montpellier	0.91 (0.48)	0.95 (0.16)	1.01 (0.08)
Clinique le Millénaire	0.88 (0.12)	0.92 (0.18)	1.00 (0.10)

5.2 Sensitivity analysis

As stated in Section 3.3, to avoid improper posterior distribution one has to assume informative prior distribution on τ_h and ω_h . Since no external prior information is available for these parameter, the definition of informative prior distribution is somehow artificial. In this section we therefore investigate if posterior inference is affected by a change in these prior distributions.

To this aim, we used more and less informative priors for τ_h and ω_h compared to that used in Section 3.3 and check whether our conclusions remains valid. Table 4 reports the evolution of the posterior means when the prior distribution for τ_h and ω_h changes to an Inverse-Gamma(3,3) to an Inverse-Gamma(1,1)—when the prior distribution is less informative. While the posterior means for $\beta_{0,h}$ and $\beta_{1,h}$ appears to be fairly robust to the prior choice, the posterior means for ω_h and τ_h appear to be strongly affected.

It is well known the the parameter τ_h and ω_h are typically negatively correlated (Zhang, 2004). For instance, a large variance τ_h combined with a large scale parameter ω_h may lead to the same overall variability than a smaller variance and small scale parameter. It is therefore more appropriate to analyse the sensitivity of the ratio τ_h/ω_h to the prior distribution definition. Table 5 shows the evolution of the posterior mean of this ratio as the prior distributions for τ_h and ω_h vary. Although at first sight some differences are noticeable, taking into account the associated uncertainties we can see that the ratio τ_h/ω_h appears to be fairly robust to the prior definition—the ratio tends to be larger as the prior distribution gets increasingly more informative though.

6 Discussion-Conclusions

In this paper, we have presented a spatial Bayesian hierarchical model based on zero inflated distribution to model count data. The two proposed models were applied to hospital admission for asthma in the Hérault department of France. To our knowledge, it is the first numerical

method to measure a distance effect on patient admissions or other similar cases, more importantly in the case of the diagnosis related groups as defined in France. For a given locality or in a relatively homogeneous area, these models can be used to compare hospital spatial attractiveness and thus to classify them from the least to the most attractive. Such ranking can be used to understand the impact of implemented reforms. Plausible ranks are obtained when comparing CHU Montpellier (first), Clinique le Millénaire (second), CH Béziers (third), and CH Bassin de Thau (least) according to the admissions due to asthma in 2009 in the Hérault. However, certain issues associated with the method are:

- Run-time: As it often the case for MCMC modules, run-times can be long. For example, for our own R software codes, with 4 hospitals and 343 localities, a chain of length 50,000 can take over 60 hours for our computer with a 2.60 GHZ Intel (R) Core (TM) i-3230M CPU and 4 Go of memory;
- Our model does not take into consideration the presence of other hospitals. The processes are only linked with the mass of extra-zeros (p) and r . It would be good to also consider covariates that distinguish hospitals such as the size or type of hospital;
- We used a linear distance function, which might be very restrictive. Further studies including both travelling time as a covariate and other functional relations such as step functions, polynomial, ... should be performed.

Compliance with Ethical Standards

We, the authors, declare that we have no conflicts of interest and that this article does not contain any studies with human participants or animals performed by any of us.

A Appendix

In this section, we give the derivation of conditional parameters distributions for the ZIP model. To obtain the derivation for the ZINB model, the distributions used in the ZIP model have to be replaced by those used in the ZINB model. As a reminder,

a) Bayes formula is

$$\pi(\theta | x) = \frac{\pi(x | \theta)\pi(\theta)}{\int \pi(x | \theta)\pi(\theta) d\theta} \propto \pi(x | \theta)\pi(\theta),$$

where $\pi(\theta | x)$ is called the posterior distribution, $\pi(\theta)$ the prior distribution and $\pi(x | \theta)$, the likelihood.

b) and the prior distributions used in this case study for the ZIP model are

$$\begin{aligned} \beta_{0,h} &\sim \frac{1}{10^3\sqrt{2\pi}} \exp\left\{-\frac{\beta_{0,h}^2}{2*10^6}\right\} & \beta_{1,h} &\sim \frac{1}{10^3\sqrt{2\pi}} \exp\left\{-\frac{\beta_{1,h}^2}{2*10^6}\right\} \\ \tau_h &\sim \frac{1}{\Gamma(1)}(1/\tau_h)^2 \exp\{-1/\tau_h\} & \omega_h &\sim \frac{1}{\Gamma(1)}(1/\omega_h)^2 \exp\{-1/\omega_h\} \\ \theta &\sim \frac{1}{10^3\sqrt{2\pi}} \exp\left\{-\frac{\theta^2}{2*10^6}\right\} \end{aligned}$$

Let's call $\pi\{y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\}$ (with $\lambda(\mathbf{x})$, β_0 , β_1 , τ and ω as defined in section 4), the joint distribution for model parameters and the data. Using the definition of conditional probabilities and assuming parameters independence within their prior distributions, we show that

$$\begin{aligned} \pi\{y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} &= \pi\{y(\mathbf{x}) | \theta, \lambda(\mathbf{x})\} \pi\{\theta\} \times \\ &\prod_{h=1}^4 \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi\{\beta_{0,h}\} \pi\{\beta_{1,h}\} \pi\{\tau_h\} \pi\{\omega_h\} \quad (4) \end{aligned}$$

Then using Bayes formula and equation 4, we derive the conditional distributions

$$\begin{aligned}
& \pi\{\theta | y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} \propto \pi\{y(\mathbf{x}) | \theta, \lambda(\mathbf{x})\} \pi(\theta) \\
& \pi\{\lambda_h(\mathbf{x}) | y(\mathbf{x}), \lambda_{-h}(\mathbf{x}), \theta, \beta_0, \beta_1, \tau, \omega\} \propto \pi\{y(\mathbf{x}) | \theta, \lambda(\mathbf{x})\} \times \\
& \quad \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \\
& \pi\{\beta_{0,h} | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} \propto \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\beta_{0,h}) \\
& \pi\{\beta_{1,h} | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} \propto \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\beta_{1,h}) \\
& \pi\{\tau_h | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} \propto \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\tau_h) \\
& \pi\{\omega_h | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} \propto \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\omega_h)
\end{aligned}$$

which have to be used in the Gibbs sampling. To obtain the explicit form of these distributions, every term in the right members has to be replaced by its mathematical expression. Thus, let's denote $n_{1,h}$ the number of locations where we count zero stays for a hospital h and $n_{2,h}$ the number of locations where we have at least one stay in the hospital.

From equation 1, subsection 3.1,

$$\begin{aligned}
\pi(y(\mathbf{x}) | \theta, \lambda(\mathbf{x})) &= \prod_{h=1}^4 \prod_{i=1}^{n_{1,h}} \left[\frac{1}{1 + \exp\{-\theta\}} + \left(1 - \frac{1}{1 + \exp\{-\theta\}}\right) \exp\{-\lambda_h(x_i)\} \right] \\
& \quad \times \prod_{j=1}^{n_{2,h}} \left[\left(1 - \frac{1}{1 + \exp\{-\theta\}}\right) \frac{\lambda_h(x_j)^{y(x_j)}}{y(x_j)!} \exp\{-\lambda_h(x_j)\} \right] \quad (5)
\end{aligned}$$

and by definition in subsection 3.2

$$\begin{aligned}
\pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} &= (2\pi)^{-n/2} |\gamma_h(\mathbf{x})|^{-1/2} \times \\
& \quad \prod_{i=1}^n \lambda_h^{-1}(x_i) \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]^T \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \quad (6)
\end{aligned}$$

By replacing every term by its explicit expression, we obtain

$$\begin{aligned}
\pi(\theta | y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega) &\propto \prod_{h=1}^4 \prod_{i=1}^{n_{1,h}} \left[\frac{1}{1 + \exp\{-\theta\}} + \frac{\exp\{-\theta\}}{1 + \exp\{-\theta\}} \exp\{-\lambda_h(x_i)\} \right] \\
& \quad \times \left[\frac{\exp\{-\theta\}}{1 + \exp\{-\theta\}} \right]^{n_{2,h}} \exp \left\{ -\frac{\theta^2}{2 * 10^6} \right\} \\
\pi(\beta_{0,h} | y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega, \theta) &\propto \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\
& \quad \times \exp \left\{ -\frac{\beta_{0,h}^2}{2 * 10^6} \right\} \\
\pi(\beta_{1,h} | y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega, \theta) &\propto \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\
& \quad \times \exp \left\{ -\frac{\beta_{1,h}^2}{2 * 10^6} \right\}
\end{aligned}$$

$$\pi(\tau_h | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau_{-h}, \omega) \propto \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\ \times |\gamma_h(\mathbf{x})|^{-1/2} (1/\tau_h)^2 \exp\{-1/\tau_h\}$$

$$\pi(\omega_h | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega_{-h}) \propto \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\ \times |\gamma_h(\mathbf{x})|^{-1/2} (1/\omega_h)^2 \exp\{-1/\omega_h\}$$

$$\pi(\lambda_h(\mathbf{x}) | y(\mathbf{x}), \theta, \lambda_{-h}(\mathbf{x}), \beta_0, \beta_1, \tau, \omega) \propto \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\ \times \prod_{k=1}^n \lambda_h^{-1}(x_k) \prod_{j=1}^{n_{2,h}} \left[\frac{\lambda_h(x_j)^{y(x_j)}}{y(x_j)!} \exp\{-\lambda_h(x_j)\} \right] \\ \times \prod_{i=1}^{n_{1,h}} \left[\frac{1}{1 + \exp\{-\theta\}} + \frac{\exp\{-\theta\}}{1 + \exp\{-\theta\}} \exp\{-\lambda_h(x_i)\} \right]$$

References

- Agarwal, D. K., Gelfand, A. E., and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological statistics*, 9(4):341–355.
- Aho, K. (2015). *asbio: A Collection of Statistical Tools for Biologists*. R package version 1.1-5.
- Banerjee, S., Carlin, B., and Gelfand, A. (2004). Hierarchical modeling and analysis for spatial statistics. *New York: Chapman&Hall/CRC*.
- Berger, J. O., De Oliveira, V., and Sansó, B. (2001). Objective bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456):1361–1374.
- Chevreur, K., Durand-Zaleski, I., Bahrami, S., Hernández-Quevedo, C., and Mladovsky, P. (2010). Health system review. *Erasmus*, 12:93–103.
- Cohen, A. C. (1963). *Estimation in mixtures of discrete distributions*. Statistical Pub. Society.
- Davison, A., Padoan, S., and Ribatet, M. (2012). Statistical modelling of spatial extremes. *Statistical Science*, 7(2):161–186.
- Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). Bayesian data analysis. texts in statistical science series.
- Greene, W. H. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models.
- Johnson, N. L. and Kotz, S. (1969). *Distributions in statistics*. J. Wiley.
- Jung, K., Feldman, R., and Scanlon, D. (2011). Where would you go for your next hospitalization? *Journal of health economics*, 30(4):832–841.

- 1
2
3
4
5
6
7
8 Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.
9
10
11 Proper, C. and Leckie, G. (2011). Increasing competition between providers in health care
12 markets: the economic evidence.
13
14 R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation
15 for Statistical Computing, Vienna, Austria.
16
17 Sang, H. and Gelfand, A. (2010). Continuous spatial process models for spatial extreme values.
18 *Journal of Agricultural, Biological, and Environmental Statistics*, 15(1):49–65.
19
20 Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures
21 of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical
22 Methodology)*, 64(4):583–639.
23
24 Varkevisser, M., van der Geest, S. A., and Schut, F. T. (2010). Assessing hospital competition
25 when prices don't matter to patients: the use of time-elasticities. *International Journal of Health
26 Care Finance and Economics*, 10(1):43–60.
27
28 Victoor, A., Delnoij, D. M., Friele, R. D., and Rademakers, J. J. (2012). Determinants of patient
29 choice of healthcare providers: a scoping review. *BMC health services research*, 12(1):272.
30
31 Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based
32 geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

CHAPITRE 4

Effet de la distance sur la probabilité de choix d'un hôpital

Entre plusieurs hôpitaux en compétition, la probabilité moyenne que les patients d'une localité choisissent tel ou tel hôpital pour admission peut être révélatrice de l'attractivité que l'hôpital présente à leurs yeux ; plus cette probabilité est grande, plus l'hôpital est attractif.

Dans ce chapitre, nous utilisons pour la modélisation des données d'admission de patients, un modèle de comptage binomial bayésien hiérarchique où la possible dépendance spatiale des données est prise en compte à travers les probabilités de choisir un hôpital dans les différentes localités. Le paramètre coefficient au niveau la variable distance sera utilisé pour comparer l'attractivité des hôpitaux et faire leur classement.

La première section donne un résumé de l'utilisation du modèle, la seconde donne l'article sur la modélisation proposée avec application sur des données de cancer de prostate dans le département du Gard et la troisième est consacrée à une annexe spécifique au chapitre.

4.1 Application de ces modèles aux données d'admission de patients

4.1.1 cadre et précisions

Supposons que nous ayons pour l'étude, le nombre de localités (indexées de $i = 1, \dots, N$) où au moins un malade nécessitant une admission a été observé, dans la zone concernée. Tout comme au chapitre précédent, les localités seront confondues à leurs coordonnées géographiques. Pour son admission, le patient peut choisir n'importe quel hôpital de la zone, hôpitaux que nous indexons aussi de $h = 1$ à

$h = H$. La seule restriction est l'attractivité (coût, qualité des soins, commodité..) de l'hôpital telle qu'il la perçoit. En ne considérant que les localités avec au moins un malade nécessitant une admission, toute zéro admission au niveau de la localité pour un hôpital suppose le manque d'attractivité de l'hôpital pour les patients de la localité.

Toujours comme au précédent chapitre, pour la distance entre localités, nous utilisons la distance euclidienne entre coordonnées géographiques de ces localités au lieu de la distance routière.

4.1.2 Modélisation des données d'admission de patients

Selon qu'ils estiment qu'un hôpital peut ou non satisfaire à leur besoin (coût, qualité des soins, commodité ...) les patients d'une localité i peuvent choisir tel ou tel hôpital h pour admission. Il est certes vrai que tous les patients n'ont pas les mêmes critères ou exigences, mais en admettant qu'il n'y a pas trop de disparité entre patients, on peut supposer qu'ils peuvent le faire avec une certaine probabilité moyenne $p_h(x_i)$. Comme tomber malade dépend de la prévalence de la maladie $I(x_i)$, la probabilité qu'un habitant de la localité i de population R_i soit admis à l'hôpital h est $p_h(x_i)I(x_i)$. Le fait d'être admis peut être vu comme la réalisation d'une loi de Bernoulli pour un résident. De ce fait, en admettant l'indépendance des admissions, le nombre de patients $y(x_i)$ est la somme des Bernoulli pour les R_i résidents, donc une loi binomiale :

$$Y(x_i) = y(x_i) \mid N(x_i), I(x_i), p_h(x_i) \sim \mathcal{B}(N(x_i), p_h(x_i) \cdot I(x_i)). \quad (4.1)$$

4.1.3 Prise en compte de la corrélation spatiale

Au chapitre précédent nous parlions déjà que des données des localités voisines sont corrélées, corrélation prise en compte au niveau des $\lambda_h(x_i)$, moyennes des Poisson. Pour les mêmes raisons, il sera plus réaliste de supposer les probabilités de choix au niveau des localités corrélées entre elles. C'est ainsi qu'après avoir paramétré $p_h(x_i)$ par

$$p_h(x_i) = \frac{1}{1 + \exp\{-\theta_h(x_i)\}},$$

la corrélation a été prise au niveau du paramètre $\theta_h(x_i)$, en supposant un processus gaussien. Ainsi en notant $\mathbf{x} = (x_1, \dots, x_n)$ et $\theta_h(\mathbf{x}) = (\theta_h(x_1), \dots, \theta_h(x_n))^t$, on a

$$\theta_h(\mathbf{x}) \sim \frac{1}{(2\pi)^{n/2} |\gamma_h(\mathbf{x})|^{1/2}} \times \exp\left[-\frac{1}{2}\{\theta_h(\mathbf{x}) - \mu_h(\mathbf{x})\}^t \gamma_h(\mathbf{x})^{-1} \{\theta_h(\mathbf{x}) - \mu_h(\mathbf{x})\}\right], \quad (4.2)$$

où $\mu_h(x_i)$ et $\gamma_h(\mathbf{x})$ sont respectivement la moyenne et la matrice de covariance du processus. Pour $\mu_h(x_i)$ tout comme pour $\gamma_h(\mathbf{x})$ plusieurs types de fonctions peuvent être utilisées. Mais pour une question de simplicité, nous avons supposé que les éléments de $\gamma_h(\mathbf{x})$ sont définis par

$$\gamma_{ij,h} = \tau_h \exp \left\{ -\frac{\|x_i - x_j\|}{\omega_h} \right\},$$

famille exponentielle isotrope et

$$\mu_h(x_i) = \beta_{0,h} + \beta_{1,h} \cdot \|x_i - x_h\|,$$

fonction linéaire de la distance entre l'hôpital et la localité.

$\beta_{1,h}$ est le paramètre qui contrôle comment varie $\mu_h(x_i)$, donc $p_h(x_i)$ au fur et à mesure qu'on s'éloigne de l'hôpital : plus il est grand, plus la probabilité serait grande, donc l'hôpital plus attractif. Il sera utilisé pour comparer l'attractivité des hôpitaux.

Une corrélation spatiale peut être utilisée pour les $I(x_i)$ mais dans notre article nous avons considéré

$$I(x_i) = \frac{A(x_i)}{R_i}.$$

Pour l'estimation des paramètres, nous avons choisi un cadre bayésien et les méthodes MCMC.

4.2 Article



MEASURING EFFECT OF DISTANCE ON PROBABILITY OF CHOOSING HOSPITAL BY PATIENTS

I. Saley¹, N. Molinari² and M. Ribatet¹

¹IMAG

UMR CNRS 5149

Université de Montpellier

CC 51, 4 Place Eugène Bataillon

34095 Montpellier Cedex 5

France

²Service DIM

CHRU de Montpellier

39 Avenue Charles Flahaut

34 295 Montpellier Cedex 5

France

Abstract

Travel distance is one of the most important drivers of patients' decision. In this work, we present a modelling of spatial number of patients' admissions for hospitals. Thereby, we numerically measure the effect of distance on the probability of choosing a hospital, allowing subsequently comparison of hospitals attractiveness related to the distance, by using a spatial Bayesian binomial hierarchical model. For estimation, we used Markov chains Monte Carlo (MCMC) methods.

Received: June 11, 2016; Accepted: August 20, 2016

Keywords and phrases: hospital attractiveness, binomial hierarchical model, Bayesian inference, MCMC methods.

1. Introduction

Factors of hospital choice include price, quality, distance or travel time, waiting time, provider network and others. Their relative importance differs according to the market characteristics and the regulatory context (Lavery et al. [9]). Early studies identified distance or travel time as the major factor negatively affecting hospital choice (Victoor et al. [12]). Sensitivity to distance varies with patients' characteristics (age, ethnicity, income and religion), admission types (i.e., stronger effect of distance for common procedures) and hospital type (Porell and Adams [10]).

Information about performance and quality of hospitals is not yet published in France. Actually, hospital choice is made by patients themselves, following advice from their general practitioners. Since general practitioners do not face any financial incentives to refer their patients to a given hospital, we can assume that they take into account patients' preferences among other factors. Thus, we can hypothesize that reputation (Jung et al. [7]), as perceived by the patients and the general practitioners is partially reflecting quality of care and easy access. In areas where there are many hospitals within short distance, attractiveness of a hospital (easy access, care quality,...) can be measured by the distance travelled by patients to get there. That is to say, longer the distance travelled, the more attractive the hospital.

In this context, a spatial modelling on hospitals choice may be very useful to propose a classification of hospitals (most to least attractive).

In this work, we use a spatial Bayesian binomial hierarchical model to model the number of patients' admissions in a given hospital according to the distance between patients and hospital locations and subsequently make hospital attractiveness comparison.

Section 2 gives the detail on the data modelling. Parameters estimation method is provided in Section 3. In Section 4, the results of an application to real data are presented. Section 5 is for short discussion and conclusion.

2. The Model

Suppose that we have n localities, indexed from 1, ..., n , where at least a sick due the pathology is observed during the period of the study in the concerned region. Suppose also no restriction is made for a sick to choose a hospital for admission apart from how he perceives its attractiveness. We mean by attractiveness, factors that include easy access, price, health care quality and so on. That is to say, any counted zero admission in a locality for a hospital in competition, in these conditions, can mean less attractiveness for the sick of the locality, compared to other hospitals where they went.

But before to give the model, let us denote

x_i : The geographic coordinates of locality indexed i and x_h , those of hospital h location. x_h can one of x_i .

$N(x_i)$: The number of residents in locality i .

$I(x_i)$: The frequency of the pathology in $N(x_i)$.

$A(x_i)$: The registered number of sick due to the pathology in locality i .

$y(x_i)$: Number of sick from locality i due to the pathology, admitted in hospital h .

The aim of the study is to find how the probability to choose a hospital h varies as one moves away the hospital and subsequently compare hospitals attractiveness by using a spatial Bayesian hierarchical model.

2.1. Data layer

Sicks from locality i can choose to go for admission in hospital h or not according to that the hospital can meet or not their need. It is true that all sick have not the same perception of the hospital, however we suppose they can choose it with a certain average probability that we note $p_h(x_i)$. Considering the fact of being sick random, we can say that the probability to be admitted for a resident of locality i in hospital h , is the probability to get sick that multiplies the probability to choose the hospital for admission.

If $I(x_i)$ is the frequency of the pathology, then a such probability is $p_h(x_i) \cdot I(x_i)$. Thus, an admitted resident of locality i is drawn from the population $N(x_i)$ with the probability $p_h(x_i) \cdot I(x_i)$. Supposing admissions independent when $p_h(x_i) \cdot I(x_i)$ known, therefore it is natural to use a binomial distribution to model the number of residents of locality i , $y(x_i)$, that have been admitted in hospital h , i.e.,

$$Y(x_i) = y(x_i) | N(x_i), I(x_i), p_h(x_i) \sim \mathcal{B}(N(x_i), p_h(x_i) \cdot I(x_i)). \quad (1)$$

2.2. Spatial correlation

In the previous subsection, instead of individual level probability of choosing a hospital, we considered an average probability to locality level; admitting implicitly common lineaments of individuals living together. So, it will also be realistic to consider that neighboring localities may share habits and information; which means the choice of ones can influence the choice of others. Thus, it will make sense to consider spatial correlation between probabilities $p_h(x_i)$, $i = 1, \dots, n$. Therefore, after we transformed $p_h(x_i)$ so as

$$p_h(x_i) = \frac{1}{1 + \exp\{-\theta_h(x_i)\}},$$

the spatial correlation is taken into account on $\theta_h(x_i)$. In such case, a Gaussian process is generally used, i.e., by noting $\mathbf{x} = (x_1, \dots, x_n)$ a $\theta_h(\mathbf{x}) = (\theta_h(x_1), \dots, \theta_h(x_n))^t$, then

$$\begin{aligned} \theta_h(\mathbf{x}) &\sim \frac{1}{(2\pi)^{n/2} |\gamma_h(\mathbf{x})|^{1/2}} \\ &\times \exp\left[-\frac{1}{2} \{\theta_h(\mathbf{x}) - \mu_h(\mathbf{x})\}^t \gamma_h(\mathbf{x})^{-1} \{\theta_h(\mathbf{x}) - \mu_h(\mathbf{x})\}\right]. \end{aligned} \quad (2)$$

We assessed that the mean of process at locality i is so as $\mu_h(x_i) = \beta_{0,h} + \beta_{1,h} \cdot \|x_i - x_h\|$; a linear function of distance between hospital and resident locations.

The parameters $\beta_{1,h}$ exhibits how the mean of θ_h varies, subsequently p_h , as one moves away the hospital h . It can be used for hospitals comparison. For the covariance matrix $\gamma_h(\mathbf{x})$, parametric structure is generally used to reduce over-parametrization. In our case, we used exponential family function, more for its simplicity than any other consideration. The elements of the matrix are so as

$$\gamma_{ij,h} = \tau_h \exp\left\{-\frac{\|x_i - x_j\|}{\omega_h}\right\}.$$

Like for $p_h(x_i)$, spatial correlation can exist between $I(x_i)$ due for example to environmental characteristics and/or people habits. A spatial structure can be used but in the present paper, we only used $\frac{A(x_i)}{N(x_i)}$ plus a constant positive number to approximate $I(x_i)$.

2.3. Priors

In Bayesian framework, unknown parameters are treated random and they are assigned distributions called *priors*. Different ways are used in setting these priors distributions according to the information about the parameters or mathematical reason. In our case we have no clear information about the parameters apart the sign of value they can take. To $\beta_{0,h}$ and $\beta_{1,h}$, we set normal distribution with mean equals to zero and variance, 10^6 , so as to be very flat in the way to be not informative. Non informative prior can be also used for τ_h and ω but some authors (Berger et al. [2]) or (Banerjee et al. [1]), showed that improper priors for these parameters can yield to improper posterior distributions. Banerjee et al. [1] suggested to use informative priors. Thus, we set to them inverse-gamma with scale and shape parameters equal to 1. This distribution has the advantage to be not far from uniform and to avoid the problem of upper bound for example for ω_h when using bounded uniform. A sensitivity analysis is recommended for these

parameters even though we did not do so since our aim is to present a methodology. To recap, parameters are set priors so as

$$\beta_{0,h} \sim \frac{1}{10^3 \sqrt{2\pi}} \exp\left\{-\frac{\beta_{0,h}^2}{2 * 10^6}\right\}, \quad \beta_{1,h} \sim \frac{1}{10^3 \sqrt{2\pi}} \exp\left\{-\frac{\beta_{1,h}^2}{2 * 10^6}\right\},$$

$$\tau_h \sim \frac{1}{\Gamma(1)} (1/\tau_h)^2 \exp\{-1/\tau_h\}, \quad \omega_h \sim \frac{1}{\Gamma(1)} (1/\omega_h)^2 \exp\{-1/\omega_h\}. \quad (3)$$

3. Estimation

To make inference, draws from $\pi\{\theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h | y(\mathbf{x})\}$, are needed.

In such situation, Gibbs sampling (Gelfand and Smith [5]), is generally used and that needs in our case conditional distributions in equation (4) which are obtained by simply using the definition of conditional probabilities and Bayes formula. But from equations (1)-(3), as one can see, these conditional distributions are of complex form. That is to say Gibbs sampling is not straightforward.

$$\begin{aligned} \pi\{\theta_h(\mathbf{x}) | y(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \\ \propto \pi\{y(\mathbf{x}) | \theta_h(\mathbf{x})\} \pi\{\theta_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}, \\ \pi\{\beta_{0,h} | y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{1,h}, \tau_h, \omega_h\} \propto \pi\{\theta_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\beta_{0,h}), \\ \pi\{\beta_{1,h} | y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \tau_h, \omega_h\} \propto \pi\{\theta_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\beta_{1,h}), \\ \pi\{\tau_h | y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \omega_h\} \propto \pi\{\theta_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\tau_h), \\ \pi\{\omega_h | y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h\} \propto \pi\{\theta_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\omega_h). \quad (4) \end{aligned}$$

So, we used Metropolis-Hasting steps (Gelman et al. [6]), within a Gibbs sampler to update each parameter. For proposal distributions, normal

distribution was used for $\theta_h(x_i)$, $\beta_{0,h}$ and $\beta_{1,h}$ and log-normal was used for τ_h and ω_h .

4. Illustration

4.1. Data

The data we used are those described in (Bounnafous and Saccareau [3]). It is about prostate cancer patients' admissions in two hospitals of Nîmes: CHU de Nîmes, public and polyclinique grand sud, private for profit. Their study listed 46 localities where at least one admission due to prostate cancer has been registered in the region. Thus, we have the number of patients in each locality and the number of patients that have been admitted in each of the two hospitals per locality. Euclidean distance from geographic coordinates of localities is used instead of road distance for the reason that we have a process with many localities and having road distance between them can be tricky.

4.2. Results

Ten MCMC chains of 100,000 iterations, are run with our codes. For each chain, we removed the draws of the 10,000 first iterations considered as burn-in period of the chain, then in the remaining iterations, we kept the draws by step of 10 iterations to reduce the dependence within draws so as to have chains of length 9,000. The computed mean and its standard error over the 10 chains, for all parameters are reported in Table 1.

These means seem to be stable chiefly for the parameter of interest, $\beta_{1,h}$. The variability observed for the means of τ_h and ω_h can be explained by the fact that their means can not be estimated consistently for a fixed domain, but only their ratio, (Zhang [13]). Thus, the draws of these chains can be used for any inference on the parameters, which means that the values of $\beta_{1,h}$ in Table 1 can be used for hospitals comparison.

Table 1. Table of computed mean of parameters and its standard error over 10 MCMC chains for the prostate cancer dataset

Parameter	CHU Nîmes	Polyclinique grand sud
$\beta_{0,h}$	-1.796 (0.178)	-1.3673 (0.0900)
$\beta_{1,h}$	-0.0165 (0.0039)	-0.0516 (0.0028)
τ_h	1.1593 (0.3225)	0.7424 (0.1820)
ω_h	11.1778 (6.3592)	8.2489 (5.6146)

5. Discussion and Conclusion

In this work, we used spatial Bayesian binomial hierarchical model to fit the data of hospitals frequentation for a given disease. It allows to measure numerically the effect of distance on patients' probability choice of hospital for admission or other similar cases and subsequently to compare hospitals attractiveness, in order, for example to understand how impacted an implemented reform.

Plausible results are obtained from the application: $\beta_{h,1}$ negative, indicating as one may expect, that distance negatively affects the choice probability and the ranks of centers. In fact, for the ranks, CHU de Nîmes, public and with more accommodation, is expected to be less affected by distance than polyclinique grand sud, private, more specialized in acute care like surgery. However, model can be improved:

(a) By using spatial correlation between the prevalences $I(x_i)$, $i = 1, \dots, n$.

(b) The process does not take into consideration the presence of other hospitals in the mean of $\theta_h(x_i)$. It considers only the distance between the hospital and patient residence locations. It may be good to consider the relative distance from other hospitals and some other covariates that can distinguish hospitals.

(c) In our modelling, we considered an average probability for given locality. It may be also interesting to consider an individual level probability which can be a local average plus a noise depending on individuals covariates such age.

References

- [1] S. Banerjee, B. Carlin and A. Gelfand, *Hierarchical Modeling and Analysis for Spatial Statistics*, New York, Chapman & Hall/CRC, 2004.
- [2] J. O. Berger, V. De Oliveira and B. Sansó, Objective Bayesian analysis of spatially correlated data, *J. Amer. Statist. Assoc.* 96(456) (2001), 1361-1374.
- [3] F. Bounnafous and M. Saccareau, *Analyse de l'attractivité des hôpitaux publics et privés sur les malades par methode mcmc*, Rapport de stage Master 1 inédit, UM2, Montpellier, France, 2012.
- [4] K. Chevreul, I. Durand-Zaleski, S. Bahrami, C. Hernández-Quevedo and P. Mladovsky, *Health System Review*, Erasmus 12 (2010), 93-103.
- [5] A. E. Gelfand and A. F. Smith, Sampling-based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.* 85(410) (1990), 398-409.
- [6] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin, *Bayesian data analysis*, Texts in Statistical Science Series, 2004.
- [7] K. Jung, R. Feldman and D. Scanlon, Where would you go for your next hospitalization? *Journal of Health Economics* 30(4) (2011), 832-841.
- [8] C. J. Lako and P. Rosenau, Demand-driven care and hospital choice, Dutch health policy toward demand-driven care: results from a survey into hospital choice, *Health Care Analysis* 17(1) (2009), 20-35.
- [9] A. A. Laverty, P. C. Smith, U. J. Pape, A. Mears, R. M. Wachter and C. Millett, High-profile investigations into hospital safety problems in England did not prompt patients to switch providers, *Health Affairs* 31(3) (2012), 593-601.
- [10] F. W. Porell and E. K. Adams, Hospital choice models: a review and assessment of their utility for policy impact analysis, *Medical Care Research and Review* 52(2) (1995), 158-195.
- [11] M. Varkevisser, S. A. van der Geest and F. T. Schut, Assessing hospital competition when prices do not matter to patients: the use of time-elasticities, *International Journal of Health Care Finance and Economics* 10(1) (2010), 43-60.

- [12] A. Victoor, D. M. Delnoij, R. D. Friele and J. J. Rademakers, Determinants of patient choice of healthcare providers: a scoping review, *BMC Health Services Research* 12(1) (2012), 272.
- [13] H. Zhang, Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics, *J. Amer. Statist. Assoc.* 99(465) (2004), 250-261.

4.3 Annexe

Dans cette section, comme nous ne l'avons pas fait dans l'article, nous donnons comment les distributions conditionnelles de l'équation (equation 4 dans l'article) :

$$\begin{aligned}
 \pi\{\theta_h(\mathbf{x}) \mid y(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} &\propto \pi\{y(\mathbf{x}) \mid \theta_h(\mathbf{x})\}\pi\{\theta_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \\
 \pi\{\beta_{0,h} \mid y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{1,h}, \tau_h, \omega_h\} &\propto \pi\{\theta_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\beta_{0,h}) \\
 \pi\{\beta_{1,h} \mid y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \tau_h, \omega_h\} &\propto \pi\{\theta_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\beta_{1,h}) \\
 \pi\{\tau_h \mid y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \omega_h\} &\propto \pi\{\theta_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\tau_h) \\
 \pi\{\omega_h \mid y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h\} &\propto \pi\{\theta_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\omega_h),
 \end{aligned} \tag{4.3}$$

pour l'échantillonneur de Gibbs sont obtenues et leur forme explicite. La hiérarchie entre les paramètres de ces lois est telle que présentée par la figure 4.1

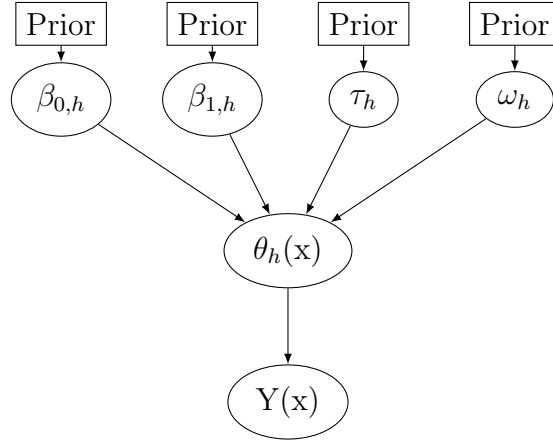


FIGURE 4.1 – Le graphe acyclique direct du modèle

Pour rappel,

a) nous avons assigné à ces paramètres des lois a priori telles que :

$$\begin{aligned}
 \beta_{0,h} &\sim \frac{1}{10^3\sqrt{2\pi}} \exp\left\{-\frac{\beta_{0,h}^2}{2 \cdot 10^6}\right\} & \beta_{1,h} &\sim \frac{1}{10^3\sqrt{2\pi}} \exp\left\{-\frac{\beta_{1,h}^2}{2 \cdot 10^6}\right\} \\
 \tau_h &\sim \frac{1}{\Gamma(1)}(1/\tau_h)^2 \exp\{-1/\tau_h\} & \omega_h &\sim \frac{1}{\Gamma(1)}(1/\omega_h)^2 \exp\{-1/\omega_h\}
 \end{aligned}$$

b) la formule de Bayes est donnée par :

$$\pi(\theta \mid x) = \frac{\pi(x \mid \theta)\pi(\theta)}{\int \pi(x \mid \theta)\pi(\theta) d\theta} \propto \pi(x \mid \theta)\pi(\theta),$$

$\pi(\theta \mid x)$, $\pi(\theta)$ et $\pi(x \mid \theta)$ étant appelées respectivement loi a priori, la loi a posteriori et la vraisemblance.

Alors soit $\pi\{y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}$, loi jointe des paramètres et observations. Par définition des probabilités conditionnelles et l'hypothèse selon laquelle ces paramètres sont indépendants à travers les lois a priori, alors on montre que

$$\begin{aligned} \pi\{y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} &= \pi\{y(\mathbf{x}) \mid \theta_h(\mathbf{x})\} \times \\ &\pi\{\theta_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi\{\beta_{0,h}\} \pi\{\beta_{1,h}\} \pi\{\tau_h\} \pi\{\omega_h\} \end{aligned} \quad (4.4)$$

En utilisant la formule de Bayes et l'équation 4.4, on déduit les formules de l'équation 4.3, qui sont utilisées pour la mise en œuvre de l'échantillonneur de Gibbs. Et pour avoir l'expression explicite de ces formules, on remplace les termes du membre droit des formules par leur expression mathématique.

De l'équation 1 de l'article, on a

$$\begin{aligned} \pi(y(\mathbf{x}) \mid \theta_h(\mathbf{x})) &= \\ \prod_{i=1}^n \binom{N(x_i)}{y(x_i)} &\left[\frac{I(x_i)}{1 + \exp\{-\theta_h(x_i)\}} \right]^{y(x_i)} \left[1 - \frac{I(x_i)}{1 + \exp\{-\theta_h(x_i)\}} \right]^{N(x_i) - y(x_i)} \end{aligned} \quad (4.5)$$

et de l'équation 2 de l'article, on a

$$\begin{aligned} \pi\{\theta_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} &= \frac{1}{(2\pi)^{n/2} |\gamma_h(\mathbf{x})|^{1/2}} \times \\ &\exp \left[-\frac{1}{2} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \}^t \gamma_h(\mathbf{x})^{-1} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \} \right] \end{aligned} \quad (4.6)$$

En remplaçant chaque terme par son expression explicite, on obtient

$$\begin{aligned} \pi(\beta_{0,h} \mid y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{1,h}, \tau_h, \omega_h) &\propto \exp \left[-\frac{1}{2} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \}^t \gamma_h(\mathbf{x})^{-1} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \} \right] \\ &\times \exp \left\{ -\frac{\beta_{0,h}^2}{2 * 10^6} \right\} \end{aligned}$$

$$\begin{aligned} \pi(\beta_{1,h} \mid y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \tau_h, \omega_h) &\propto \exp \left[-\frac{1}{2} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \}^t \gamma_h(\mathbf{x})^{-1} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \} \right] \\ &\times \exp \left\{ -\frac{\beta_{1,h}^2}{2 * 10^6} \right\} \end{aligned}$$

$$\begin{aligned} \pi(\tau_h \mid y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \omega_h) &\propto \exp \left[-\frac{1}{2} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \}^t \gamma_h(\mathbf{x})^{-1} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \} \right] \\ &\times |\gamma_h(\mathbf{x})|^{-1/2} (1/\tau_h)^2 \exp\{-1/\tau_h\} \end{aligned}$$

$$\begin{aligned} \pi(\omega_h \mid y(\mathbf{x}), \theta_h(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h) &\propto \exp \left[-\frac{1}{2} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \}^t \gamma_h(\mathbf{x})^{-1} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \} \right] \\ &\times |\gamma_h(\mathbf{x})|^{-1/2} (1/\omega_h)^2 \exp \{ -1/\omega_h \} \end{aligned}$$

$$\begin{aligned} \pi(\theta_h(\mathbf{x}) \mid y(\mathbf{x}), \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h) &\propto \exp \left[-\frac{1}{2} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \}^t \gamma_h(\mathbf{x})^{-1} \{ \theta_h(\mathbf{x}) - \mu_h(\mathbf{x}) \} \right] \times \\ &\prod_{i=1}^n \left[\frac{I(x_i)}{1 + \exp\{-\theta_h(x_i)\}} \right]^{y(x_i)} \left[1 - \frac{I(x_i)}{1 + \exp\{-\theta_h(x_i)\}} \right]^{N(x_i) - y(x_i)} \end{aligned}$$

CQFD.

Envergure territoriale d'un hôpital

L'objectif de ce chapitre est de donner une méthode pour mesurer l'envergure territoriale d'un hôpital, c'est à dire la distance maximale que pourraient accepter de parcourir des patients pour se rendre dans cet hôpital donné ; et subséquemment de classer les hôpitaux en compétition du plus attractif au moins attractif selon cette envergure : un hôpital avec une grande envergure est plus attractif que celui avec une envergure moins grande.

Pour ce faire, nous considérons l'hôpital comme un produit. Son attractivité peut être vue comme l'utilité qu'il présente aux yeux des patients en fonction de la distance qu'ils parcourent pour s'y rendre. En pareil cas, des modèles de choix discret comme les modèles d'Utilité Aléatoire (Random Utility Models ou RUMs en abrégé) peuvent être utilisés.

Une application du modèle Logit est présentée à la section 5.1. Mais compte tenu de certaines limites constatées au vu de notre objectif, nous avons introduit une nouvelle approche concernant le choix qu'opèrent les individus. La section 5.2 présente un résumé de la nouvelle approche et un article utilisant cette approche et soumis pour évaluation à International Journal of Health Economics and Management est donné à la section 5.3.

5.1 Application d'un RUMs : le Logit

L'application porte sur des données d'admission due à l'asthme en 2009 dans le département de Pyrénées-Orientales (Languedoc-Roussillon). Il s'agit plus précisément du nombre des patients ayant été admis dans deux des hôpitaux (Table 5.1) de ce département : CH Perpignan situé à Perpignan et Polyclinique Saint Roch à Cabestany et dont nous voulons comparer l'attractivité aux yeux des patients.

Nous avons les localisations (coordonnées géographiques) des patients et des hôpitaux, donc la distance euclidienne entre le lieu de résidence du patient et la localité abritant l'hôpital.

TABLE 5.1 – Nombre de séjours enregistrés dus à l'asthme par chacun des deux hôpitaux en 2009

Hôpital	Nombre de séjours enregistrés
CH Perpignan	212
Polyclinique Saint Roch	39

5.1.1 modèle

Le but de l'étude est de mesurer l'impact de la distance sur le choix des patients en ce qui concerne les deux hôpitaux évoqués ci-haut pour séjour hospitalier, en utilisant un modèle de choix discret. C'est dire que nous considérons les hôpitaux comme des produits parmi lesquels devrait choisir un patient pour son séjour. Nous pouvons supposer que le choix opéré par le patient peut se ramener à la satisfaction à certain degré de son besoin, autrement dit, à l'utilité que présente l'hôpital à ses yeux en tenant compte de certaines variables qui dépendent de lui et/ou de l'hôpital ; une situation de choix où nous pouvons utiliser les RUMs. Donc si nous considérons un patient n et un hôpital j , l'utilité que présente cet hôpital à ses yeux peut être écrite comme $U_{nj} = V_{nj} + \epsilon_{nj}$ où V_{nj} est la partie de l'utilité expliquée par des variables observables et ϵ_{nj} , la partie inconnue traitée comme aléatoire par l'analyste. Et dans cette étude, nous avons supposé que la seule variable observable influençant le choix du patient est la distance qui sépare sa localité de résidence à celle abritant l'hôpital. De ce fait, nous pouvons écrire que

$$V_{nj} = \alpha_j + \beta_j \|x_n - x_j\|$$

où α_j et β_j sont des constantes spécifiques à l'hôpital et $\|x_n - x_j\|$, la distance entre la localité de résidence du patient et celle abritant l'hôpital. Donc

$$U_{nj} = \alpha_j + \beta_j \|x_n - x_j\| + \epsilon_{nj}.$$

Comme nous sommes avec des alternatives où aucune raison de corrélation entre ϵ_{n1} et ϵ_{n2} n'est évidente, alors nous pouvons aussi supposer que ϵ_{n1} et ϵ_{n2} sont indépendants et utiliser le modèle Logit, c'est-à-dire que la distribution des erreurs est

$$f(\epsilon_{nj}/\sigma) = e^{-\epsilon_{nj}/\sigma} e^{-e^{-\epsilon_{nj}/\sigma}},$$

où σ^2 est la variance des ϵ_{nj} .

Pour rappel, la probabilité de choix d'une alternative avec le modèle Logit est

$$P_{ni} = \frac{e^{V_{ni}/\sigma}}{\sum_j e^{V_{nj}/\sigma}}$$

et de façon explicite pour notre cas d'étude,

$$P_{ni} = \frac{e^{\alpha_i/\sigma + \beta_i \|x_n - x_i\|/\sigma}}{\sum_j e^{\alpha_j/\sigma + \beta_j \|x_n - x_j\|/\sigma}}.$$

Et la probabilité du choix d'un patient n peut être écrite telle que

$$\prod_{i=1}^2 (P_{ni})^{y_{ni}} = \prod_{i=1}^2 \left(\frac{e^{\alpha_i + \beta_i \|x_n - x_i\|}}{\sum_j e^{\alpha_j + \beta_j \|x_n - x_j\|}} \right)^{y_{ni}}$$

où $y_{ni} = 1$ si le patient n a choisi l'hôpital i et 0 sinon. En assumant le choix des patients indépendants entre eux, la vraisemblance serait

$$LH = \prod_{n=1}^{286} \prod_{i=1}^2 \left(\frac{e^{\alpha_i + \beta_i \|x_n - x_i\|}}{\sum_j e^{\alpha_j + \beta_j \|x_n - x_j\|}} \right)^{y_{ni}}.$$

Mais σ n'étant pas connu et n'ayant ni incidence sur la probabilité de choix ni sur la comparaison des paramètres, dorénavant nous l'omettrons pour simplicité tout en gardant à l'esprit que les paramètres qui seront estimés sont en réalité les paramètres de V_{nj} divisés par ce paramètre.

5.1.2 Estimation des paramètres

La méthode de maximum de vraisemblance est généralement utilisée pour l'estimation des paramètres mais nous, nous la ferons dans le cadre bayésien pour la raison que tous nos autres travaux ont été faits dans un cadre bayésien. Ce qui nous conduit à une réécriture de la vraisemblance. Considérons A comme l'ensemble des choix des hôpitaux pour admission tels que opérés par les patients. Nous pouvons dire que la probabilité d'obtenir cet ensemble sachant les paramètres est

$$P(A \mid \alpha_i, \beta_i, i = 1, 2) = \prod_{n=1}^{286} \prod_{i=1}^2 \left(\frac{e^{\alpha_i + \beta_i \|x_n - x_i\|}}{\sum_j e^{\alpha_j + \beta_j \|x_n - x_j\|}} \right)^{y_{ni}}.$$

Pour ces modèles, les constantes α_i n'étant pas identifiables mais leur différence, il y a lieu de fixer arbitrairement une de ces constantes. C'est ainsi que nous avons fixé α_1 à zéro. Ainsi on aura la vraisemblance telle que

$$P(A \mid \alpha_i^*, \beta_i, i = 1, \dots, 2) = \prod_{n=1}^{286} \prod_{i=1}^2 \left(\frac{e^{\alpha_i + \beta_i \|x_n - x_i\|}}{\sum_j e^{\alpha_j + \beta_j \|x_n - x_j\|}} \right)^{y_{ni}}$$

où $\alpha_i^* = \alpha_i - \alpha_1$. Avec deux alternatives, on a qu'une seule constante, α_2^* qui reste inconnue car α_1^* étant égale à zéro. En inférence bayésienne, des lois a priori sont à assigner aux paramètres. Comme chacun de nos paramètres peut être positif ou négatif, nous leur avons assigné chacun une loi normale de moyenne 0 et de variance 10^6 pour être proche d'une uniforme, donc non informative.

En utilisant la formule de Bayes, on a

$$P(\alpha_2^*, \beta_i, i = 1, 2 \mid A) \propto \prod_{n=1}^{286} \prod_{i=1}^2 \left(\frac{e^{\alpha_2^* + \beta_i \|x_n - x_i\|}}{\sum_j e^{\alpha_j^* + \beta_j \|x_n - x_j\|}} \right)^{y_{ni}} \pi(\alpha_2^*, \beta_i, i = 1, 2)$$

qui est la loi a posteriori dont nous avons besoin des réalisations pour faire de l'inférence sur les paramètres. Pour en avoir pour ce genre de distributions, des méthodes Monte-Carlo par Chaines de Markov sont généralement utilisées ; et plus précisément, l'échantillonneur de Gibbs [Gelfand and Smith, 1990]. Pour le mettre en œuvre, les lois conditionnelles (obtenues par simple application des probabilités conditionnelles et la formule de Bayes en supposant les paramètres indépendants à travers leur lois a priori) données dans l'équation 5.1 sont utilisées. Leurs écritures explicites sont de formes complexes et peuvent ne pas être directement simulables. Alors pour la mise à jour des paramètres à chaque itération de l'échantillonneur de Gibbs, l'algorithme de Hasings-Metropolis [Gelman et al., 2004] est mis à contribution avec des lois normales comme lois de proposition pour chaque paramètre.

$$\begin{aligned} \pi(\alpha_2^* \mid A, \beta_1, \beta_2) &\propto \prod_{n=1}^{286} \prod_{i=1}^2 \left(\frac{e^{\alpha_2^* + \beta_i \|x_n - x_i\|}}{\sum_j e^{\alpha_j^* + \beta_j \|x_n - x_j\|}} \right)^{y_{ni}} e^{-\frac{(\alpha_2^*)^2}{2 \times 10^6}} \\ \pi(\beta_1 \mid A, \alpha_2^*, \beta_2) &\propto \prod_{n=1}^{286} \prod_{i=1}^2 \left(\frac{e^{\alpha_2^* + \beta_i \|x_n - x_i\|}}{\sum_j e^{\alpha_j^* + \beta_j \|x_n - x_j\|}} \right)^{y_{ni}} e^{-\frac{(\beta_1)^2}{2 \times 10^6}} \\ \pi(\beta_2 \mid A, \alpha_2^*, \beta_1) &\propto \prod_{n=1}^{286} \prod_{i=1}^2 \left(\frac{e^{\alpha_2^* + \beta_i \|x_n - x_i\|}}{\sum_j e^{\alpha_j^* + \beta_j \|x_n - x_j\|}} \right)^{y_{ni}} e^{-\frac{(\beta_2)^2}{2 \times 10^6}} \end{aligned} \quad (5.1)$$

5.1.3 résultats

Pour collecter des réalisations de la loi a posteriori, une chaîne de 500.000 itérations est lancée avec les codes MCMC que nous avons programmés sous le logiciel R [R Core Team, 2017]. Les 100.000 premières itérations de la chaîne, considérées comme temps de chauffe de la chaîne, ont été supprimées. Puis les éléments de la chaîne ont été gardés par pas de 10 pour réduire la dépendance à travers ces éléments pour ainsi avoir finalement une chaîne de longueur 40.000. Pour vérifier la convergence de la chaîne, nous avons tracé des graphes et des distributions conditionnelles des paramètres. Les figures 5.1 et 5.2 présentent à la fois le graphe de la chaîne et la distribution conditionnelle pour le paramètre β_j respectivement le CH Perpignan et la Polyclinique Saint Roch, qui semblent montrer une bonne convergence des paramètres. La table 5.2 reporte la moyenne de chaque paramètre estimée et son intervalle de crédibilité à 95%. Et de ces résultats, le CH Perpignan semble des deux, le moins affecté par la distance ; situation, somme toute logique car le CH est un établissement public plus adapté pour ce genre de soins que la Polyclinique qui est un établissement privé offrant moins de capacité et plus spécialisé dans les soins plus aigus.

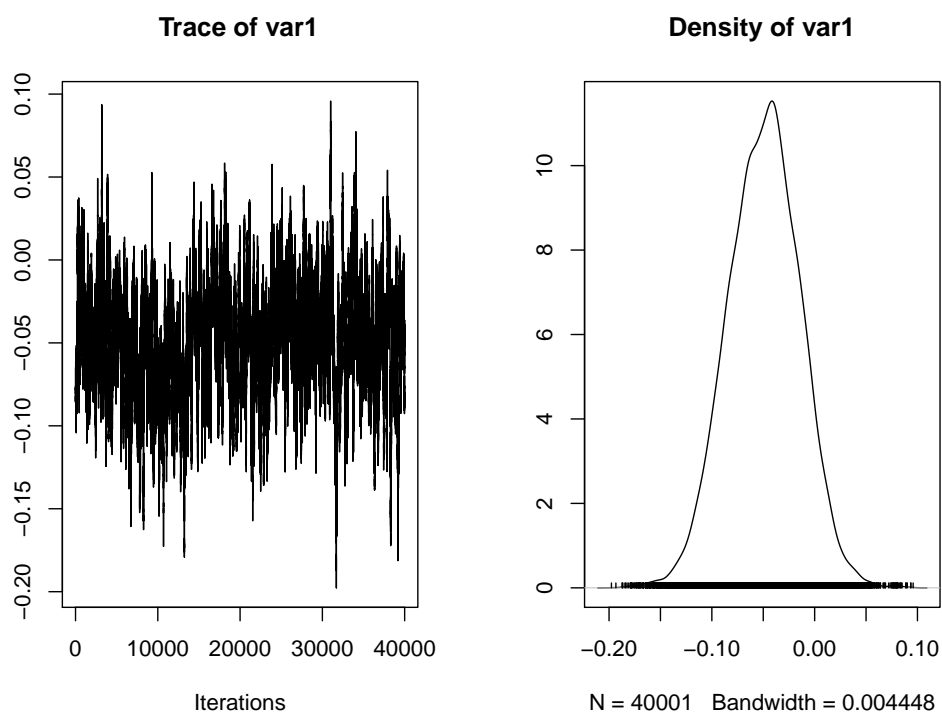


FIGURE 5.1 – Chaîne MCMC et densité conditionnelle du paramètre effet distance pour le CH de Perpignan

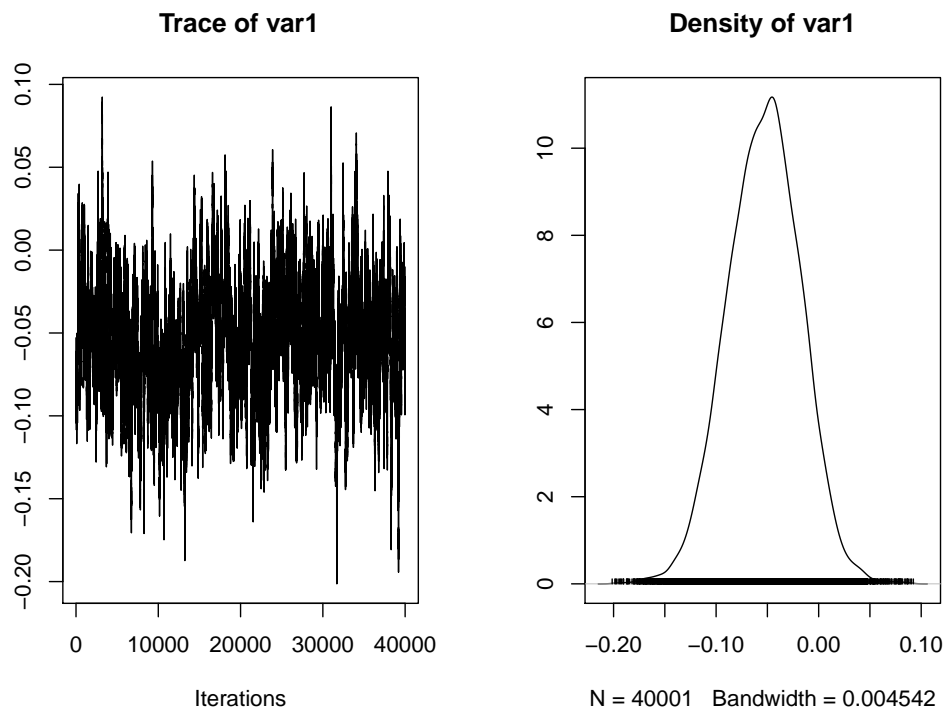


FIGURE 5.2 – Chaîne MCMC et densité conditionnelle du paramètre effet distance pour la Polyclinique Saint-Roch

TABLE 5.2 – Table des valeurs moyennes des β_j et leur intervalle de crédibilité à 95%

Hôpital	La moyenne de β_j	IC à 95%
CH Perpignan	-0.0493	(-0.1178 ; 0.0165)
Polyclinique Saint Roch	-0.0533	(-0.1232 ; 0.0140)

5.1.4 Limites des modèles RUMs au vu de notre but

Certes le modèle a donné des résultats plausibles concernant l'impact de la distance sur les deux centres. En fait, vu que c'est des données d'asthme, on s'attendrait à ce que CH Perpignan qui est public et plus grand soit plus adapté que la Polyclinique Saint-Roch qui est privée et plus orientée dans des soins tels que la chirurgie. Nous aurions voulu étendre cette application sur plusieurs hôpitaux. Mais malheureusement nous ne pouvons l'appliquer que si tous les hôpitaux sont dans des villes distinctes. Sinon, du fait que nous avons considéré que les coordonnées géographiques des lieux de résidence des patients et les localisations des hôpitaux, pour deux hôpitaux situés dans une même localité, β_j ne peut être identifiable. Aussi, plus que l'effet-distance sur l'hôpital, notre objectif était de mesurer la distance au delà de laquelle, un patient, en ne considérant que l'aspect distance, ne choisirait l'hôpital en question pour admission. Par exemple en supposant l'utilité représentative, V_{nj} , positive et décroissante au fur et à mesure qu'on s'éloigne de l'hôpital, (c'est-à-dire α_j positif et β_j négatif), on peut admettre qu'un patient ne considérant que la distance, ne peut choisir un hôpital j dès lors que $V_{nj} \leq 0$. Cette distance limite d_j est telle que

$$-\frac{1}{d_j} = r_j$$

où

$$r_j = \frac{\beta_j}{\alpha_j}.$$

Mais malheureusement pour les RUMs, α_j ne peut être estimé du fait que la probabilité du choix est basée sur la différence des utilités. Pour rappel, la probabilité de choix est donnée par

$$\begin{aligned} P_{nj} &= Prob(U_{nj} > U_{ni}, \forall i \neq j) \\ &= Prob(V_{nj} + \epsilon_{nj} > V_{ni} + \epsilon_{ni}, \forall i \neq j) \\ &= Prob(\epsilon_{nj} - \epsilon_{ni} > V_{ni} - V_{nj}, \forall i \neq j) \\ &= Prob(\epsilon_{nj} - \epsilon_{ni} > \alpha_i + \beta_i \|x_n - x_i\| - \alpha_j - \beta_j \|x_n - x_j\|, \forall i \neq j). \end{aligned}$$

Donc le ratio r_j ne peut être déterminé et subséquemment la distance d_j . En plus, les RUMs supposent que le client choisit le produit avec la plus grande utilité. Cette hypothèse a été beaucoup critiquée par des psychologues tels que Tversky [Tversky, 1972] et l'économiste Simon (rationalité limitée). Il nous semble aussi difficile qu'un malade choisisse toujours un hôpital avec la plus grande utilité pour

lui. Pour pallier à l'estimation de d_j et prendre en compte le cas où le patient ne maximise pas l'utilité, nous avons introduit une nouvelle approche.

5.2 Une autre approche pour calculer l'envergure

5.2.1 L'approche proposée

Dans cette approche, on relâche l'hypothèse selon laquelle un agent choisira le produit qui lui procure le plus d'utilité pour une plus souple : un agent peut choisir un produit dès lors que ce dernier lui procure une certaine utilité qu'il juge satisfaisante mais non nécessairement maximale, en considérant certains aspects. Nous justifions cette hypothèse d'une part, par le fait que comme le disait Tversky [Tversky, 1972], un agent devant choisir entre plusieurs alternatives n'est souvent pas sûr de savoir laquelle choisir ou ne ferait pas toujours le même choix dans des situations identiques. Dès lors, il est fort probable que l'agent ne maximise pas toujours l'utilité pour choisir. D'autre part, d'après Simon (rationalité limitée), un agent ne considère que certains aspects du produit. En plus, il est peu probable qu'un agent choisisse une alternative qui ne lui serait d'aucune utilité ou plutôt lui nuirait. Donc nous pouvons supposer qu'il faut qu'il ait un minimum d'utilité positive pour choisir une alternative en considérant certains aspects. C'est ce minimum d'utilité que nous appelons utilité-seuil.

Cette utilité-seuil peut varier d'un individu à un autre du fait de leurs réalités socio-démographiques et d'autres facteurs inconnus. Elle peut varier pour un même individu en fonction du temps et de ses conditions socio-démographiques. Mais comme nous ne l'observons pas, nous la considérons aléatoire et lui assignons une distribution. Ainsi la probabilité de toute alternative d'être choisie peut être calculée.

Pour fixer les idées, considérons qu'un agent n d'une utilité-seuil S_n doit choisir entre J alternatives, sans qu'elles soient nécessairement mutuellement exclusives. Soit U_{nj} , l'utilité (déterministe en considérant que certains aspects de l'alternative) que lui procurerait l'alternative j . Il peut choisir j dès lors que $U_{nj} \geq S_n$ et nous avons défini la probabilité de choix de j comme

$$P_{nj} = Pr(U_{nj} \geq S_n), \quad (5.2)$$

probabilité qui suppose aussi que plus l'utilité U_{nj} est grande, plus j a la chance d'être choisie. Ainsi l'alternative qui procure la plus grande utilité à l'agent a plus de chance d'être choisie sans qu'elle ne soit obligatoirement choisie.

Du fait qu'un agent ne choisirait ni une alternative d'aucune utilité pour lui, ni celle qui lui nuirait, nous pouvons considérer son utilité-seuil S_n comme positive. Comme cette utilité-seuil peut être une conjonction de beaucoup facteurs (socio-démographiques, inconnus...) qui peuvent être indépendants entre eux, nous pouvons

utiliser une loi log-normale pour la modéliser, c'est-à-dire :

$$f(S_n = s) = \frac{1}{s\sigma_n\sqrt{2\pi}} \exp\left\{-\frac{(\log s - \mu_n)^2}{2\sigma_n^2}\right\}, \quad s > 0. \quad (5.3)$$

Ou encore

$$P_{nj} = Pr(U_{nj} \geq S_n) = F_{\mu_n, \sigma_n}(U_{nj}), \quad (5.4)$$

où F_{μ_n, σ_n} est la fonction de répartition de la loi log-normale de paramètres (μ_n, σ_n) . Mais μ_n ne peut être connue du fait que l'utilité absolue n'est pas mesurable. Ce qui nous amène, en utilisant la propriété que pour tout x ,

$$F_{\mu_n, \sigma_n}(x) = F_{0, \sigma_n}\left(\frac{x}{\exp(\mu_n)}\right)$$

pour une distribution log-normale, à réécrire la probabilité de choix de telle sorte que

$$P_{nj} = Pr(U_{nj} \geq S_n) = F_{\mu_n, \sigma_n}(U_{nj}) = F_{0, \sigma_n}(U_{nj}^r), \quad (5.5)$$

avec $U_{nj}^r = U_{nj} / \exp(\mu_n)$, une utilité relative à estimer.

La probabilité de choix d'un agent peut être écrite comme

$$\prod_j \{P_{nj}\}^{y_{nj}} = \prod_j \{F_{0, \sigma_n}(U_{nj}^r)\}^{y_{nj}}, \quad (5.6)$$

où $y_{nj}=1$ si l'alternative j est choisie par l'agent n and 0 sinon.

Pour un échantillon de N agents devant choisir parmi les J alternatives et considérant les S_n indépendants, la vraisemblance de A , l'ensemble des choix des alternatives tels que observés, est

$$LH(A) = \prod_{n=1}^N \prod_j \{P_{nj}\}^{y_{nj}}. \quad (5.7)$$

5.2.2 Application aux données d'admission de patients

Considérons que nous avons N patients dans une région, chacun devant choisir un hôpital pour admission parmi J hôpitaux de la région. Nous supposons que le seul aspect observable qui détermine leur choix est la variable distance qu'ils doivent parcourir pour se rendre à l'hôpital. De ce fait, nous définissons l'utilité qu'obtiendrait un patient n en choisissant un hôpital j comme

$$U_{nj} = \alpha_j^* + \beta_j^* \|x_n - x_j\|$$

où α_j^* est un paramètre positif exprimant l'utilité de l'hôpital à domicile, x_n, x_j , les coordonnées géographiques de la localité de résidence du patient n et celle de l'hôpital d'admission et $\|x_n - x_j\|$, la distance euclidienne entre ces localités. Le paramètre β_j^* de l'effet de la distance. Comme logiquement l'attractivité d'un hôpital

diminue au fur et à mesure qu'on s'en éloigne, le paramètre β_j^* est nécessairement négatif. Étant donné l'hypothèse que l'utilité-seuil S_n est distribuée log-normale, alors pour $U_{nj} \leq 0$, la probabilité qu'un patient n , choisisse l'hôpital j est nulle. Ce qui veut dire, tout patient habitant au delà de la distance telle que $\alpha_j^* + \beta_j^* \cdot d_j = 0$, ne choisirait l'hôpital j pour admission. C'est cette distance que nous appelons envergure territoriale de l'hôpital en question. Plus elle est grande, plus l'hôpital est attractif. Elle sera utilisée pour comparer l'attractivité des hôpitaux. Elle est subséquemment estimée en estimant le ratio défini par

$$r_j = \frac{\beta_j^*}{\alpha_j^*}.$$

Pour ce faire, notons A , l'ensemble des choix des hôpitaux pour admission tels que opérés par les patients. Pour faire simple dans notre cas et éviter une sur-paramétrisation du modèle, nous avons supposé que les paramètres (μ_n, σ_n) sont les mêmes pour tous les patients de la région, c'est-à-dire que $(\mu_n, \sigma_n) = (\mu, \sigma), \forall n$, en considérant que dans une zone où il n'y a pas trop de disparité, les habitants ont pratiquement les mêmes attentes. Ainsi nous pouvons définir l'utilité relative U_{nj}^r par

$$U_{nj}^r = \frac{\alpha_j^*}{\exp(\mu)} + \frac{\beta_j^*}{\exp(\mu)} \cdot \|x_n - x_j\| = \alpha_j + \beta_j \cdot \|x_n - x_j\|,$$

définition qui est sans conséquence sur r_j car

$$\frac{\beta_j}{\alpha_j} = \frac{\beta_j^*}{\exp(\mu)} \cdot \frac{\exp(\mu)}{\alpha_j^*} = \frac{\beta_j^*}{\alpha_j^*} = r_j.$$

En supposant S_n , ($n = 1, \dots, N$) indépendants, alors la probabilité d'observer l'ensemble des choix A , sachant σ , α_j , β_j , $j = 1, \dots, J$ peut s'écrire comme

$$\pi(A \mid \alpha_1, \dots, \alpha_J, \beta_1, \dots, \beta_J, \sigma) = \prod_{n=1}^N \prod_j \{P_{nj}\}^{y_{nj}} \quad (5.8)$$

où $P_{nj} = F_{0,\sigma}(\alpha_j + \beta_j \cdot \|x_n - x_j\|)$ et $F_{0,\sigma}$, la fonction de répartition de la loi log-normale de paramètres $(0, \sigma)$. $y_{nj} = 1$ quand le patient n a choisi l'hôpital j et 0 sinon.

Pour l'estimation des paramètres, le maximum de vraisemblance peut être utilisé. Cependant, dans notre travail nous nous sommes placé dans un contexte d'estimation bayésienne par souci de cohérence avec les travaux présentés dans les chapitres précédents.

5.3 Article

International Journal of Health Economics and Management
**Measuring hospital spatial wingspan by using a discrete choice model with utility-
threshold.**
--Manuscript Draft--

Manuscript Number:	IHFE-D-17-00040
Full Title:	Measuring hospital spatial wingspan by using a discrete choice model with utility-threshold.
Article Type:	Research article
Keywords:	Discrete choice models; Hospital Attractiveness; Bayesian Inference; MCMC Methods
Corresponding Author:	Issa SALEY Université de Montpellier Montpellier, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Université de Montpellier
Corresponding Author's Secondary Institution:	
First Author:	Issa SALEY
First Author Secondary Information:	
Order of Authors:	Issa SALEY
	Mathieu Ribatet
	Nicolas Molinari
Order of Authors Secondary Information:	
Funding Information:	

Measuring hospital spatial wingspan by using a discrete choice model with utility-threshold

Issa Saley · Mathieu Ribatet · Nicolas Molinari

Received: date / Accepted: date

Abstract Policy makers increasingly rely on hospital competition to incentivize patients to choose high-value care. Travel distance is one of the most important drivers of patients' decision. The paper presents a method to numerically measure, for a given hospital, the distance beyond which no patient is expected to choose the hospital for treatment by using a new approach in discrete choice models. Thereby, we compared 3 hospitals attractiveness related to this distance for asthma patients admissions in 2009 in Hérault (France), showing, as expected, CHU Montpellier is the one with the most important spatial wingspan. For estimation, Monte Carlo Markov Chain (MCMC) methods are used.

Keywords Discrete choice models · Hospital Attractiveness · Bayesian Inference · MCMC Methods.

I. Saley

IMAG, UMR CNRS 5149, Université de Montpellier, CC 51, 4 place Eugène Bataillon, 3409 Montpellier Cedex 5, France E-mail: saleyissa@yahoo.fr

M. Ribatet

IMAG, UMR CNRS 5149, Université de Montpellier, CC 51, 4 place Eugène Bataillon, 3409 Montpellier Cedex 5, France

N. Molinari

IMAG, UMR CNRS 5149, Université de Montpellier, CC 51, 4 place Eugène Bataillon, 3409 Montpellier Cedex 5, France

N. Molinari

Service DIM, CHRU de Montpellier, 39 avenue Charles Flahaut, 34 295 Montpellier Cedex 5, France

Measuring hospital spatial wingspan by using a discrete choice model with utility-threshold.

•

Received: date / Accepted: date

Abstract Policy makers increasingly rely on hospital competition to incentivize patients to choose high-value care. Travel distance is one of the most important drivers of patients' decision. The paper presents a method to numerically measure, for a given hospital, the distance beyond which no patient is expected to choose the hospital for treatment by using a new approach in discrete choice models. Thereby, we compared 3 hospitals attractiveness related to this distance for asthma patients admissions in 2009 in Hérault (France), showing, as expected, CHU Montpellier is the one with the most important spatial wingspan. For estimation, Monte Carlo Markov Chain (MCMC) methods are used.

Keywords Discrete choice models · Hospital Attractiveness · Bayesian Inference · MCMC Methods.

1 Introduction

It is often advocated that competition between hospitals may improve efficiency and quality of health care [15]. In fact, in theory, due to increasing competition, hospitals decrease production costs and improve the quality of health care delivered so as to attract new patients. Factors of hospital choice include price, quality, distance or travel time (easy access), waiting time, provider network and others. Their relative importance differs according to the market characteristics and the regulatory context [10]. Early studies identified distance or travel time as the major factor negatively affecting hospital choice [19]. In addition, patients frequently cite distance among the factors influencing their choice [9]. Sensitivity to distance varies with patient characteristics (age, ethnicity, income and religion), admission types (i.e., stronger effect of distance

Address(es) of author(s) should be given

1 for common procedures) and hospital type (i.e., weaker effect for admissions
2 to larger specialized hospitals) [14].

3
4 As in many European countries, French hospitals are financed through
5 Diagnosis Related Groups based on a prospective payment system. Regarding
6 acute care, this scheme has been fully implemented in 2005 for private for-
7 profit hospital budgets and in 2008 for public hospital budgets. The reform
8 was intended to improve efficiency and fairness in financing and also to increase
9 competition between and within the public and private sectors [5]. Information
10 about performance and quality of hospitals is not yet published in France,
11 and waiting times are currently not identified as an issue. Actually, hospital
12 choice is made by patients themselves, following advice from their general
13 practitioners. Since general practitioners do not face any financial incentives
14 to refer their patients to a given hospital, we can assume that they take into
15 account patients' preferences among other factors. Thus, we can hypothesize
16 that reputation [8], as perceived by the patients and the general practitioners is
17 partially reflecting the quality of care and easy access. In these circumstances,
18 travel distance certainly plays a key role in patients' decision [18]. Further, in
19 areas where there are many hospitals within short distance, the attractiveness
20 of a hospital (easy access, care quality, . . .) can be measured by the distance
21 travelled by patients to get there. That is to say, longer the distance travelled,
22 the more attractive the hospital.

23 In this context, a spatial modeling on hospitals choice may be very useful
24 to propose a classification of hospitals (most to least attractive). From a policy
25 perspective, these findings have implications for regulation. For instance, the
26 effect of making the quality of care information publicly available may not have
27 the expected impact in areas where facilities are scarce and distant from one to
28 another. More generally, the predominant effect of distance on hospitals choice
29 should be acknowledged and taken into consideration before implementing any
30 reforms aimed at incentivizing hospitals competitiveness.

31 In this work, we present a method to numerically measure the distance
32 beyond which no patient is expected to choose the hospital for treatment
33 by using a new approach in discrete choice models. This distance is used to
34 compare hospitals attractiveness.

35 For illustration, we used patients admissions for asthma in 3 hospitals of
36 Hérault that have been registered in 2009.

37 The paper is organized as follows. Section 2 describes the data we used
38 for illustration while Section 3 explains the new approach. The illustration is
39 presented in section 4 and the paper ends with a brief discussion.

40 41 42 43 44 **2 Data**

45 The data we used in this work come from the *Programme de Médicalisation*
46 *des Systèmes d'Information* (PMSI). The PMSI is an administrative dataset
47 recording all inpatient admissions in French private and public hospitals cover-
48 ing all social health insurance schemes. It is intended to describe the medical
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

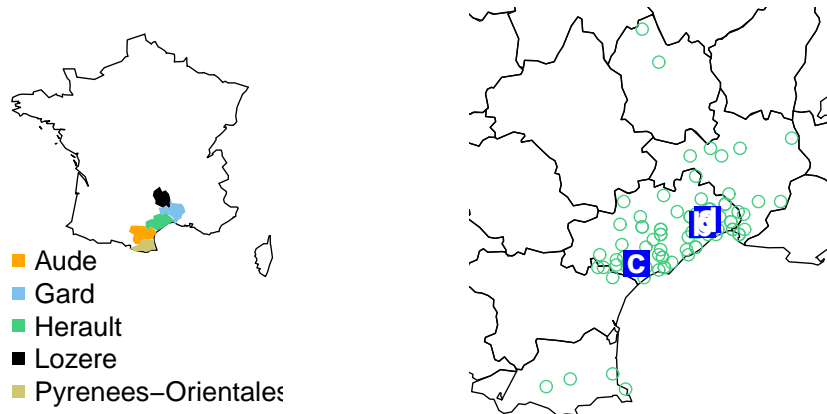


Fig. 1 Mapping of the illustration data. The left map shows where the Region Languedoc-Roussillon is located in France. The picture on the right displays locations of the hospitals (a: CHU Montpellier, b: Clinique le Millénaire and c: CH Béziers) and localities (green points) where at least one resident admission is reported in these hospitals

activities of hospitals. Information about the patient and the hospital of stay such as the patient disease, his place of residence, location of the hospital, type of hospital, are held in the dataset.

For our case study, we used data about asthma hospital patients stays in Languedoc-Roussillon, a region known to have a high competitive health care market in France. Located in the South of France, see Figure 1, the region consists of five departments: Aude, Gard, Hérault, Lozère and Pyrénées-Orientales. It extends from 42° to 44.5° Latitude North and between 1.5° to 5° Longitude East. The Languedoc-Roussillon region has a population of about 3 million people, which includes both rural and metropolitan areas. It is the second region with the highest population growth (1.1%) in France (about 0.4%) due essentially to immigrants and more than 16% of the population are retirees.

In 2009, 1,289 hospital patient stays have been recorded for asthma in this region as reported in [1]. The study focused on admissions in 3 hospitals of Hérault: one is private for-profit and 2 are not-for-profit. Table 1 displays the hospitals with the number of stays.

Table 1 Number of asthma stays registered per each of the 3 hospitals in 2009

Hospital	number of asthma patient stays
CH Béziers	79
CHU Montpellier	127
Clinique Le Millénaire	23

For each of these 3 hospitals, we extracted from the dataset their geographic coordinates, the number of patients admitted per locality and the geographic coordinates of these localities. From the geographic coordinates, we computed euclidean distance between the hospital and patient locations, used as travel distance, because having road distance between localities can be tricky.

3 Proposed approach

3.1 Motivation

Patients admissions to hospitals can be seen as a situation of choice where agents (here patients) have to choose among a set of products or alternatives (here hospitals for a given therapy); therefore discrete choice models can be used to model the data. Since it seems reasonable to consider that a patient chooses an hospital when obtaining a certain utility of doing so, Random Utility models (RUMs) [12] can be used.

The concept of these models is that for any given alternative j from J alternatives in competition, an agent n is expected to obtain a certain utility U_{nj} of choosing the alternative. This utility U_{nj} is often decomposed in the sum of 2 parts:

- a) V_{nj} (called representative utility), the part of utility described by the observed variables influencing the choice of the agent;
- b) ϵ_{nj} , the part of utility related to unobserved variables also influencing the choice of the agent,

i.e, $U_{nj} = V_{nj} + \epsilon_{nj}$. By considering ϵ_{nj} , $j = 1, \dots, J$, random, thus assigning them a distribution f and under the assumption that the agent chooses the alternative with the greatest utility, the probability of choosing an alternative j by the agent n can be derived as

$$\begin{aligned}
 P_{nj} &= \Pr(U_{nj} > U_{ni}, \forall i \neq j) \\
 &= \Pr(V_{nj} + \epsilon_{nj} > V_{ni} + \epsilon_{ni}, \forall i \neq j) \\
 &= \Pr(\epsilon_{nj} - \epsilon_{ni} > V_{ni} - V_{nj}, \forall i \neq j)
 \end{aligned} \tag{1}$$

According to the specification of $f(\cdot)$, different specific models are obtained. For example, the Logit model [11] is obtained under the assumption that ϵ_{nj} are independent copies of a Gumbel distribution, i.e.,

$$f(\epsilon_{nj}) = e^{-\epsilon_{nj}} e^{-e^{-\epsilon_{nj}}}. \quad (2)$$

This model is by far the most widely used among RUMs due to some of its properties. But through the model use, limitations arise: independence of unknown parts may not always hold; Logit can not handle some cases of taste variation or some panel data. Various models have been developed to cope with these limitations. We can cite the Generalized Extreme Value distribution among which the most well-known is the Nested-Logit model [2]. It is derived under the assumption that the cumulative joint distribution of ϵ_{nj} , $j = 1, \dots, J$ is

$$F(\epsilon_{n1}, \dots, \epsilon_{nJ}) = \exp \left\{ - \sum_{k=1}^K \left(\sum_{i \in B_k} e^{-\epsilon_{ni}/\lambda_k} \right)^{\lambda_k} \right\}, \quad (3)$$

where B_k is a subset of the J alternatives partitioned in K non-overlapping subsets called Nests so as for 2 alternatives j and m in a same Nest B_k , ϵ_{nj} and ϵ_{nm} are correlated and for 2 alternatives j and m in 2 distinct Nests, respectively B_k and B_l , $k \neq l$, $cov(\epsilon_{nj}, \epsilon_{nm}) = 0$. The parameter λ_k measures the degree of independence of the ϵ_{nj} in Nest B_k ; the lower the value of λ_k , the more correlated the ϵ_{nj} and the higher its value, the less correlated the unobserved factors. The value of $\lambda_k = 1$ corresponds to Logit model.

Probit ([16] for binary case) is developed under the assumption that

$$f(\epsilon_{n1}, \dots, \epsilon_{nJ}) = (2\pi)^{-J/2} |\Sigma|^{-1/2} \exp \left(-\frac{1}{2} \epsilon_n^T \Sigma^{-1} \epsilon_n \right),$$

with $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nJ})^T$ and Σ the covariance matrix. The problem with this model is that the probability of choice is not closed-form and hence, requires simulations.

Mixed Logit [3] and [4], more flexible [13], is so as if β is the unknown parameter of V_{nj} , then the choice probability is

$$P_{nj} = \int_{\beta} \frac{e^{V_{nj}(\beta)}}{\sum_j e^{V_{nj}(\beta)}} g(\beta) d\beta \quad (4)$$

where g can be any density function on β .

In this study, we have patients admissions for asthma in 3 hospitals of Hérault that have been registered in 2009. To use RUMs, we define the utility that may obtain a patient n from choosing hospital j as

$$U_{nj} = \alpha_j + \beta_j \|x_n - x_j\| + \epsilon_{nj}, \quad (5)$$

that is to say

$$V_{nj} = \alpha_j + \beta_j \|x_n - x_j\|. \quad (6)$$

α_j and β_j are unknown parameters specific to the hospital j and $\|x_n - x_j\|$, the euclidean distance between patient n and hospital j locations. β_j controls how V_{nj} varies according to the distance. By assessing α_j positive and utility decreasing (thus β_j negative) by far one moves away the hospital j , the ratio

$$r_j = \frac{\beta_j}{\alpha_j} = -\frac{1}{d_j},$$

where d_j is the distance beyond which $V_{nj} \leq 0$. This distance d_j called wingspan, can be interpreted as the distance beyond which no patient is expected to choose the hospital considering the distance since the representative utility is negative.

However, first, r_j can not be estimated with RUMs. In fact, the constants α_j can not be estimated in the expression

$$\Pr(\alpha_j + \beta_j \|x_n - x_j\| - \alpha_i - \beta_i \|x_n - x_i\| > \epsilon_{ni} - \epsilon_{nj}, \forall i \neq j).$$

Only the difference $\alpha_j - \alpha_i$ can be estimated; and there is an infinity of constants α_j and α_i giving the same difference. One more problem in estimating of r_j , is that for 2 hospitals j and i in a same locality, β_j and β_i can not be estimated but their difference because $\|x_n - x_j\| = \|x_n - x_i\|$. For instance, it is the case for CHU Montpellier and Clinique le Millénaire in our study, which are both located in Montpellier. These two issues make that r_j can not be estimated with the RUMs, though knowing this ratio is needed in this study in order to compare the attractiveness of the hospitals.

Second, the assumption of utility-maximization in choosing a hospital for treatment can be not true for some patients (bounded rationality).

So, to estimate r_j and cope with the situation where patients do not maximize the utility, we propose a new approach. In our approach, the utility is really considered deterministic and the assumption of utility maximization, which is criticized by many psychologists is released for a more loose assumption by introducing a utility-threshold: any alternative whose utility for the agent n exceeds a certain threshold S_n , can be chosen by this agent.

3.2 The approach

In order to cope with the estimation issue of ratio mentioned above, we present another approach where the assumption of utility maximization used to derive the choice probability is released. In fact, when an agent is faced with a choice among several alternatives, he is often not sure which alternative he should select, and does not always make the same choice under seemingly identical conditions, as noted by Tversky [17]. Therefore, it is strongly possible that an agent does not always maximize the utility when choosing an alternative. As an agent does not always choose the alternative with the greatest utility for him and as it is unlikely he chooses an alternative that he thinks useless or harmful for him, we can suppose that he chooses an alternative from the

moment the utility he may obtain considering some aspects covers his certain need (Bounded rationality). It is this certain need that we call a threshold of utility. It is a minimum of utility that may demand the agent from an alternative. It can vary due to socio-demographic variables, the type of alternatives, and other unknown factors. Since we don't know how exactly it varies, though that these socio-demographic variables can make an agent to be more or less demanding, we treat it random and assign it a distribution. Thus, any alternative probability choice of the agent, considering some aspects can be derived. The greater the utility of the alternative is, compared to the threshold more likely the alternative is going to be chosen by the agent.

To formalize the idea, consider that we have an agent n facing choice among J alternatives. As for random utility models, these alternatives are in finite number. They do not have to be mutually exclusive, but just exhaustive in the sense that any choice of the agent may be met by the combination of these alternatives. Let U_{nj} be the utility that may obtain the agent n considering some aspects from choosing the alternative j . The agent can choose the alternative if only the utility he may obtain from doing so, exceeds a minimum of utility (called utility-threshold) S_n , i.e, $U_{nj} > S_n$. The S_n is treated as random and assigned a probability density function f . The probability of an alternative j to be chosen by the agent n is defined as

$$P_{nj} = \Pr(U_{nj} \geq S_n) \quad (7)$$

The greater U_{nj} is, compared to S_n , the more likely the alternative is going to be chosen. Therefore, the alternative with the greatest utility considering the aspects has the most chance to be chosen.

Considering that it is more realistic for an agent to choose an alternative if having a benefit, the threshold is expected to be positive. We assess that the logarithm of S_n has a normal distribution with parameters (μ_n, σ_n) , i.e,

$$f(S_n = s) = \frac{1}{s\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log s - \mu_n)^2}{2\sigma_n^2}\right\}, \quad s > 0. \quad (8)$$

Thus,

$$P_{nj} = \Pr(U_{nj} \geq S_n) = F_{\mu_n, \sigma_n}(U_{nj}), \quad (9)$$

where F_{μ_n, σ_n} is the cumulative log-normal distribution with parameters (μ_n, σ_n) . But in general, it is known that absolute utility can not be measured, so μ may not be known. We know that from the constraint, $\exp(\mu_n)$ is from zero to infinity. So we divide U_n by $\exp(\mu_n)$ and re-state the probability choice so as

$$P_{nj} = \Pr(U_{nj} \geq S_n) = F_{\mu_n, \sigma_n}(U_{nj}) = F_{0, \sigma_n}(U_{nj}^r), \quad (10)$$

where $U_{nj}^r = U_{nj} / \exp(\mu_n)$, a relative utility which is to be estimated. For equation (10), in fact, it is easy to show that for any x

$$F_{\mu, \sigma}(x) = F_{0, \sigma}\left(\frac{x}{\exp(\mu)}\right)$$

for a log-normal distribution.

The probability of an agent choice can be written as

$$\prod_j \{P_{nj}\}^{y_{nj}} = \prod_j \{F_{0,\sigma_n}(U_{nj}^r)\}^{y_{nj}} \quad (11)$$

where $y_{nj}=1$ if j is chosen by n and 0 otherwise.

For a sample of N agents choosing among J alternatives and considering that the S_n are independent, the likelihood for A , the set of observed choices, is

$$LH(A) = \prod_{n=1}^N \prod_j \{P_{nj}\}^{y_{nj}} \quad (12)$$

For unknown parameters estimation, likelihood maximization can be used; but in this paper, a Bayesian framework is adopted.

4 Illustration

Recall that the aim of this study is to compare the attractiveness of the hospitals according to the travel distance made by patients. In fact, we have the number of patients by locality, hospitals where they went and euclidean distances obtained from geographic coordinates of the localities.

4.1 Model under the proposed approach

Consider that we have N patients who face choice among J hospitals. We suppose that the only observed variable that may determine their choice is travel distance. Therefore, we define the utility that may obtain the patient n considering the distance from choosing hospital j as

$$U_{nj} = \alpha_j^* + \beta_j^* \|x_n - x_j\|$$

where α_j^* is positive, x_n, x_j , respectively geographic coordinates of the patient n and hospital j locations and $\|x_n - x_j\|$, the euclidean distance between these locations. The parameter β_j^* exhibits the effect of distance. Since the utility of a hospital is supposed to reduce as one moves away, the parameter β_j^* is expected to be negative. From the assumption that the threshold S_n is distributed log-normal, then for $U_{nj} \leq 0$, the probability that the patient n chooses the hospital j is zero. It means that beyond the distance d_j so as $\alpha_j^* + \beta_j^* \cdot d_j = 0$, no patient is supposed to choose the hospital j for treatment. This distance will be used to compare hospitals attractiveness. The bigger this distance is, the more attractive the hospital is. By estimating the ratio

$$r_j = \frac{\beta_j^*}{\alpha_j^*},$$

we can subsequently estimate d_j .

The main target is now to estimate the ratio r_j for the 3 hospitals with 229 patients stays (Table 1) by using the approach above in a Bayesian framework.

Denote A , the set of choices actually observed. To reduce over parameterization, we assume that the parameters of the log-normal distribution are the same for all the patients since people in a same community share almost the same characteristics, i.e., $(\mu_n, \sigma_n) = (\mu, \sigma), \forall n$. Then we define

$$U_{nj}^r = \frac{\alpha_j^*}{\exp(\mu)} + \frac{\beta_j^*}{\exp(\mu)} \cdot \|x_n - x_j\| = \alpha_j + \beta_j \cdot \|x_n - x_j\|,$$

the relative utility, without impact on r_j because

$$\frac{\beta_j}{\alpha_j} = \frac{\beta_j^*}{\exp(\mu)} \cdot \frac{\exp(\mu)}{\alpha_j^*} = \frac{\beta_j^*}{\alpha_j^*} = r_j.$$

Supposing that $S_n, n = 1, \dots, 229$ are independent, then the probability to have A , the set of the observed choices, when, $\sigma, \alpha_j, \beta_j, j = 1, 2, 3$ are known can be expressed as

$$\pi(A | \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \sigma) = \prod_{n=1}^{229} \prod_j \{P_{nj}\}^{y_{nj}} \quad (13)$$

where $P_{nj} = F_{0,\sigma}(\alpha_j + \beta_j \cdot \|x_n - x_j\|)$ and $F_{0,\sigma}$, cumulative log-normal distribution with parameters $(0, \sigma)$. $y_{nj} = 1$ when the agent n chooses the alternative j and zero otherwise.

It can be good to use joint distribution for some of the parameters to take into account the presence of other alternatives but we suppose them all independent. Hence, for each of α_j, β_j and σ , we set a prior distribution. For α_j , a uniform distribution can be used. But since it is unlikely for α_j to be infinite and we don't know its exact upper bound, thus, we prefer to use an Inverse-Gamma distribution with scale and shape parameters equal to 1. Also an Inverse-Gamma is used for σ but the scale and shape parameters equal 2 and for β_j , a Uniform $U(-2, 0)$, as prior distributions.

4.2 Estimation methods

For simplicity, denote $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, $\beta = (\beta_1, \beta_2, \beta_3)$, α_{-j} , α without the component α_j and β_{-j} , β without the component β_j .

To compute r_j , a sample of the parameters has to be drawn from the posterior distribution $\pi(\alpha, \beta, \sigma | A)$.

MCMC algorithms such as Gibbs sampling [6] are often used to have approximate draws from this kind of distributions. For Gibbs sampling, parameters conditional distributions given in (14) whose detailed expressions are given in Appendix (section 6) are to be used. But, these distributions are not closed-form; which means that drawing from these distributions is not straightforward.

$$\begin{aligned}
&\pi\{\alpha_j \mid A, \alpha_{-j}, \beta, \sigma\} \propto \pi\{A \mid \alpha, \beta, \sigma\} \pi(\alpha_j) \\
&\pi\{\beta_j \mid A, \alpha, \beta_{-j}, \sigma\} \propto \pi\{A \mid \alpha, \beta, \sigma\} \pi(\beta_j) \\
&\pi\{\sigma \mid A, \alpha, \beta\} \propto \pi\{A \mid \alpha, \beta, \sigma\} \pi(\sigma)
\end{aligned} \tag{14}$$

So, we used Metropolis-Hasting steps [7] within a Gibbs sampler to update each parameter. For proposal ditribution, normal distribution was used for β_j , and a log-normal was used for α_j and σ , for $j=1,2,3$.

4.3 Results

To collect sample draws, we ran 100,000 iterations with the MCMC codes. We removed the first 20,000 iterations considered as burn-in period of the chain, then in the remaining iterations, we kept the draws by step of 10 iterations to reduce the dependence within draws. To check the convergence of the MCMC chains obtained, we made plots of these chains for each parameter and traced the corresponding conditional distribution. For Clinique le Millénaire, the chains of α_j and β_j are variable and not mixing well. It is likely due to the lack of enough locations of observations: in fact, there are 13 localities without zero admission. So, we produced 10 MCMC chains and though the variability of α_j and β_j of Clinique le Millénaire, the computed mean of r_j is stable for all the 3 hospitals as reported in Table 2; which means that this ratio can be used for comparison. The table 3 gives the computed average distance at which the utility provided by the hospital is zero; which means that beyond this distance, no asthma sick is expected to come in the hospital according to the data. For recall this distance d_j is obtained by equaling $\alpha_j^* + \beta_j^*.d_j$ to zero, that's to say $1 + r_j.d_j = 0$.

Table 2 Table of the computed mean of r_j and its Standard Error (SE) over 10 MCMC chains for the asthma dataset

Hospital	r_j mean value	SE
CH Béziers	-0.0042	0.0004
CHU Montpellier	-0.0024	0.0001
Clinique le Millénaire	-0.0166	0.0003

Table 3 Table of average distance and its Standard Error (SE) over 10 MCMC chains. Beyond this distance no patient is expected to come in the hospital for treatment due to asthma

Hospital	d_j (km)	Std. Err.
CH Béziers	239	17
CHU Montpellier	426	17
Clinique le Millénaire	60.4	1.2

The results clearly show that the CHU Montpellier is the most attractive and with an important wingspan that goes beyond the Languedoc-Roussillon, the administrative zone of the center, even if this wingspan has to be relativized due to the distance used and the assumptions made on the sample. Its attractiveness is not surprising. CHU Montpellier is a well-known center providing a high level healthcare service in Languedoc Roussillon and has big accommodation capacity. It has also easy access due to transportation facilities: Highway A9, railways (TGV and TER) and many other roads. The rank of CH Béziers is plausible because it is in reality the second health center in Hérault with lower activity than CHU Montpellier.

Clinique Le Millénaire is most negatively affected by distance and with a wingspan that does not go beyond the department of Hérault. The rank of Clinique le Millénaire can be explained by the fact it is a private hospital, more specialized in acute health care like surgery and with less accommodation capacity. The results are in adequacy with what we expected for the attractiveness of the hospitals in competition.

Though the results obtained seem to be plausible, in Bayesian analysis, it is important to know how sensitive they are to the choice of the prior distributions. In our illustration, it is good to know how the r_j is sensitive to the choice of the different priors; mainly the choice of the priors of α_j and β_j . So, to check the sensitivity, we keep one of 2 priors constant and make vary the hyperparameters of the other. For β_j , we used $\text{unif}(-5,0)$, $\text{unif}(-10,0)$ and $\text{unif}(-20,0)$ and no sensible variation are observed. But when we make vary the prior of α_j , the mean of α_j becomes smaller when its prior is more informative and bigger in the opposite. And by the way, it changes the mean of β_j . Even if the change is not very big, the mean of r_j seems to become bigger for CH Montpellier and smaller for Clinique le Millénaire (which has a small number of admissions), when the prior of α_j is more informative as reported in Table 4. But when using less informative prior for α_j , we noticed that the ratio for CHU Montpellier not stable. So, caution has to be used when choosing the prior of α_j .

Table 4 Table of the computed mean of r_j over 10 MCMC chains according to the Inverse-Gamma (IG) used as distribution of α_j

	$IG(2, 2)$	$IG(3, 3)$	$IG(5, 5)$
CH Béziers	-0.004	-0.004	-0.004
CHU Montpellier	-0.002	-0.002	-0.002
Clinique le Millénaire	-0.017	-0.017	-0.017

5 Discussions-Conclusions

In this paper, we presented a method to measure hospital spatial attractiveness by using a new approach in discrete choice models. This approach releases the

assumption of utility-maximization which seems sometimes not realistic for a more loose assumption. The assumption supposes a threshold of utility above which an agent can choose an alternative. Application of the approach is done to asthma patients admissions in 2009 for 3 hospital of Hérault. Plausible results are obtained even if they have to be relativized to the assumptions made on the sample.

But, in spite of the obtained plausible results, application to other discrete choice data has to be led to further study of the approach and comparison with other models in order to know its power and limitations. For instance, utility correlation between alternatives (in a hierarchical model) can be considered to improve the model.

6 Appendix

Conditional distributions derivation

In this section, we give the derivation of parameters conditional distributions used in the Gibbs sampler (see section 4). Then the explicit form of the distributions is obtained by replacing every term of these distributions by its mathematical expression.

For recall, Bayes formula is

$$\pi(\theta | x) = \frac{\pi(x | \theta)\pi(\theta)}{\int \pi(x | \theta)\pi(\theta) d\theta} \propto \pi(x | \theta)\pi(\theta),$$

where $\pi(\theta | x)$ is called posterior distribution, $\pi(\theta)$ prior distribution and $\pi(x | \theta)$, the likelihood.

Consider $\pi(A, \alpha, \beta, \sigma)$ as the joint distribution of A , α , β (as defined in section 4) and σ .

Using the definition of conditional probabilities and assumption of independence of the parameters within their prior distributions, we show that

$$\pi(A, \alpha, \beta, \sigma) = \pi(A | \alpha, \beta, \sigma)\pi(\alpha_1) \dots \pi(\alpha_3)\pi(\beta_1) \dots \pi(\beta_3)\pi(\sigma) \quad (15)$$

Then using Bayes formula and the equation (15), we obtain

$$\pi(\alpha_j | A, \alpha_{-j}, \beta, \sigma) \propto \pi(A | \alpha, \beta, \sigma)\pi(\alpha_j)$$

$$\pi(\beta_j | A, \alpha, \beta_{-j}, \sigma) \propto \pi(A | \alpha, \beta, \sigma)\pi(\beta_j)$$

$$\pi(\sigma | A, \alpha, \beta) \propto \pi(A | \alpha, \beta, \sigma)\pi(\sigma)$$

which are the conditional distributions to be used in the Gibbs sampling. To have an explicit form of these distributions, every term in the right members has to be replaced by its mathematical expressions which are:

a) for prior distributions,

$$\pi(\alpha_j) = \frac{1}{\Gamma(1)}(1/\alpha_j)^2 \exp(-1/\alpha_j)$$

$$\pi(\beta_j) = \frac{1}{2}I_{[-2,0]}$$

$$\pi(\sigma) = \frac{4}{\Gamma(2)}(1/\sigma)^3 \exp(-2/\sigma)$$

b) and for the likelihood,

$$\pi(A | \alpha, \beta, \sigma) = \prod_{n=1}^{229} \prod_j \left[\int_0^{\alpha_j + \beta_j \cdot \|x_n - x_n\|} \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y)^2}{2\sigma^2}\right\} dy \right]^{y_{nj}}$$

where $\int_0^{\alpha_j + \beta_j \cdot \|x_n - x_n\|} \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y)^2}{2\sigma^2}\right\} dy = F_{0,\sigma}(\alpha_j + \beta_j \|x_n - x_j\|)$.

Therefore, we obtain

$$\pi(\alpha_j | A, \alpha_{-j}, \beta_1, \sigma) \propto \prod_{n=1}^{229} \prod_j \left[\int_0^{\alpha_j + \beta_j \cdot \|x_n - x_n\|} \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y)^2}{2\sigma^2}\right\} dy \right]^{y_{nj}} \frac{1}{\Gamma(1)}(1/\alpha_j)^2 \exp(-1/\alpha_j)$$

$$\pi(\beta_j | A, \alpha, \beta_{-j}, \sigma) \propto \prod_{n=1}^{229} \prod_j \left[\int_0^{\alpha_j + \beta_j \cdot \|x_n - x_n\|} \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y)^2}{2\sigma^2}\right\} dy \right]^{y_{nj}} I_{[-2,0]}$$

$$\pi(\sigma | A, \alpha, \beta) \propto \prod_{n=1}^{229} \prod_j \left[\int_0^{\alpha_j + \beta_j \cdot \|x_n - x_n\|} \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y)^2}{2\sigma^2}\right\} dy \right]^{y_{nj}} \frac{4}{\Gamma(2)}(1/\sigma)^3 \exp(-2/\sigma)$$

7 Compliance with Ethical Standards

We, the authors, declare that we have no conflict of interest and this article does not contain any studies with human participants or animals performed by any of us.

References

1. Aubas, C., Bourdin, A., Aubas, P., Gamez, A., Halimi, L., Vachier, I., Malafaye, N., Chanez, P., Molinari, N.: Role of comorbid conditions in asthma hospitalizations in the south of france. *Allergy* **68**(5), 637–643 (2013)
2. Ben-Akiva, M.: The structure of travel demand models, transportation systems division, department of civil engineering (1972)
3. Boyd, J.H., Mellman, R.E.: The effect of fuel economy standards on the us automotive market: an hedonic demand analysis. *Transportation Research Part A: General* **14**(5-6), 367–378 (1980)
4. Cardell, N.S., Dunbar, F.C.: Measuring the societal impacts of automobile downsizing. *Transportation Research Part A: General* **14**(5-6), 423–434 (1980)

-
5. Chevreul, K., Durand-Zaleski, I., Bahrami, S., Hernández-Quevedo, C., Mladovsky, P.: Health system review. *Erasmus* **12**, 93–103 (2010)
 6. Gelfand, A.E., Smith, A.F.: Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association* **85**(410), 398–409 (1990)
 7. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian data analysis. texts in statistical science series (2004)
 8. Jung, K., Feldman, R., Scanlon, D.: Where would you go for your next hospitalization? *Journal of health economics* **30**(4), 832–841 (2011)
 9. Lako, C.J., Rosenau, P.: Demand-driven care and hospital choice. dutch health policy toward demand-driven care: results from a survey into hospital choice. *Health Care Analysis* **17**(1), 20–35 (2009)
 10. Laverty, A.A., Smith, P.C., Pape, U.J., Mears, A., Wachter, R.M., Millett, C.: High-profile investigations into hospital safety problems in england did not prompt patients to switch providers. *Health Affairs* **31**(3), 593–601 (2012)
 11. Luce, R.D., Suppes, P.: Preference, utility, and subjective probability. Wiley (1965)
 12. Marschak, J.: Stanford symposium on mathematical methods in the social sciences (ed.) k arrow (1960)
 13. McFadden, D., Train, K., et al.: Mixed mnl models for discrete response. *Journal of applied Econometrics* **15**(5), 447–470 (2000)
 14. Porell, F.W., Adams, E.K.: Hospital choice models: a review and assessment of their utility for policy impact analysis. *Medical Care Research and Review* **52**(2), 158–195 (1995)
 15. Propper, C., Leckie, G.: Increasing competition between providers in health care markets: the economic evidence (2011)
 16. Thurstone, L.L.: A law of comparative judgment. *Psychological review* **34**(4), 273 (1927)
 17. Tversky, A.: Elimination by aspects: A theory of choice. *Psychological review* **79**(4), 281 (1972)
 18. Varkevisser, M., van der Geest, S.A., Schut, F.T.: Assessing hospital competition when prices don't matter to patients: the use of time-elasticities. *International Journal of Health Care Finance and Economics* **10**(1), 43–60 (2010)
 19. Victoor, A., Delnoij, D.M., Friele, R.D., Rademakers, J.J.: Determinants of patient choice of healthcare providers: a scoping review. *BMC health services research* **12**(1), 272 (2012)

CHAPITRE 6

Conclusion et Perspectives

Dans notre travail de recherche, nous avons présenté plusieurs manières de modéliser des données d'attractivité hospitalière selon l'angle que l'on considère pour l'attractivité; donnant à l'analyste un éventail de choix selon ses objectifs. Des données réelles ont été utilisées pour illustration ayant donné de plausibles résultats. Grâce à ces modélisations, l'analyste pourra prévoir le flux des patients ou classer les hôpitaux afin par exemple de définir des hôpitaux de proximité et de pointes. Mais comme toute modélisation, chacun de ces modèles a ses faiblesses.

Le premier modèle ZIM utilise les données de comptage avec excès de zéros. Il convient plus à une situation où l'analyste dans une zone déjà prédéfinie (district sanitaire, département, région...), veut savoir comment évolue le nombre moyen de patients admis dans tel ou tel hôpital au fur et à mesure qu'on s'en éloigne et éventuellement comparer les hôpitaux par l'effet de la distance sur ce nombre. Son problème est qu'il ne prend en compte que la dimension spatiale pour le nombre d'admissions en méconnaissant par exemple la présence d'autres hôpitaux. Il serait amélioré en utilisant d'autres covariables que ce soit au niveau de la masse des zéros en excès et/ou au niveau de la moyenne du processus et en supposant d'autres formes fonctionnelles pour la moyenne du processus. Nous savons aussi supposé un processus continu. Or les localités (ici communes) sont des points précis et non continus; des processus tels que les champs de Markov peuvent être utilisés.

Le modèle binomial hiérarchique est plus centré sur la probabilité de choisir un hôpital dans une localité donnée, donnant à l'analyste une autre alternative dans son étude. Il peut s'utiliser pour la comparaison de l'attractivité des hôpitaux tout comme pour calculer le nombre de patients attendus. Dans notre travail, nous avons considéré une probabilité moyenne au niveau de la localité mais il serait très intéressant de considérer la probabilité au niveau de l'individu en introduisant des covariables qui peuvent être propres à l'individu et utiliser une corrélation des pré-

valences au niveau des localités. C'est un modèle qui peut être lourd à mettre en œuvre car plus d'information à chercher : prévalence, population... Des champs de Markov aussi peuvent être utilisés.

Une autre manière de voir les données d'attractivité hospitalière, c'est de considérer les hôpitaux comme des produits, ayant pour clients, les patients. On pourra alors utiliser des modèles de choix discret ; plus précisément les modèles d'utilité aléatoire. C'est des modèles plus centrés sur l'individu dont on cherche à décrire le choix et qui prennent en compte la présence de tous les hôpitaux car basés sur la comparaison de ces hôpitaux en supposant que le client choisit le produit avec la plus grande utilité pour lui. Mais du fait de certaines limitations dans l'estimation des paramètres et pour être plus réaliste, l'hypothèse selon laquelle un individu choisit un produit avec la plus grande utilité est relâchée pour une hypothèse plus souple ; une fois que l'individu pense qu'un produit le satisfait d'une certaine manière en considérant certains aspects, il peut le choisir sans que soit la plus grande satisfaction. Avec cette approche, on peut calculer la distance au delà de laquelle, un patient ne peut choisir un hôpital ; on peut donc par exemple définir des hôpitaux de proximité et de pointe. Cette hypothèse qui nous a permis de calculer l'envergure est plus générale que celle de la maximisation de l'utilité même si pour le moment nous n'avons pas encore établi de lien mathématique entre les modèles obtenus sous les deux hypothèses ; en effet la maximisation de l'utilité en est un cas particulier. Aussi le modèle obtenu peut être vu comme un modèle à règle de décision aléatoire dès lors que l'individu ne choisit pas forcément l'alternative avec la plus utilité même si cette dernière a quand même la plus grande probabilité d'être choisie. Notre article est initiateur et il serait intéressant d'améliorer le modèle, d'en établir un lien mathématique avec les autres modèles de choix et de comparer leur performance.

Mémento sur les chaînes de Markov

Dans ce mémento, nous ne parlerons que des chaînes de Markov à temps discret à valeurs dans un espace mesurable d'une manière sommaire en donnant quelques définitions et résultats utiles (sans démonstration) en s'inspirant du chapitre 3 du livre de Robert [Robert, 1996]. Il s'agit principalement des chaînes de Markov homogènes, irréductibles, apériodiques, récurrentes et ergodiques, généralement utilisées par les algorithmes MCMC. Pour plus de détails, le lecteur intéressé pourra se référer à ce livre et surtout à [Meyn and Tweedie, 1993] comme l'a recommandé Robert lui-même.

A.1 Généralités

Soit $(\mathcal{X}, \mathcal{B})$, un espace mesurable. Une suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans \mathcal{X} (appelé espace d'états) est une **chaîne de Markov** si la loi conditionnelle de X_n sachant X_{n-1}, X_{n-2}, \dots est la même que celle de X_n sachant X_{n-1} . Autrement dit l'état futur ne dépend des états passés que par l'état présent.

Sa loi est complètement déterminée par la donnée de la loi μ de X_0 , dite loi initiale de la chaîne et de la loi conditionnelle de X_n sachant X_{n-1} , appelée **transition**.

On appelle **noyau de transition** sur un espace mesurable $(\mathcal{X}, \mathcal{B})$, toute application

$$P : \mathcal{X} \times \mathcal{B} \rightarrow R_+$$

telle que pour tout $A \in \mathcal{B}$, $P(\cdot, A)$ est mesurable et pour tout $x \in \mathcal{X}$, $P(x, \cdot)$ est une mesure de probabilité.

Par extension la densité de $P(x, \cdot)$ est également appelée noyau. Il détermine les propriétés de la chaîne par les relations suivantes :

$$P_x(X_1 \in A_1) = P(x, A_1),$$

$$P_x((X_1, X_2) \in A_1 \times A_2) = \int_{A_1} P(y_1, A_2) P(x, dy_1)$$

...

$$P_x((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) = \int_{A_1} \dots \int_{A_{n-1}} P(y_{n-1}, A_n) \\ \times P(x, dy_1) \dots P(y_{n-2}, dy_{n-1})$$

En posant $P^1(x, A) = P(x, A)$, on définit le noyau de n transitions par

$$P^n(x, A) = \int_{\mathcal{X}} P^{n-1}(y, A) P(x, dy), \quad n \geq 2.$$

Et plus généralement, on a pour tout $(m, k) \in \mathbf{N}^2, x \in \mathcal{X}, A \in \mathcal{B}$,

$$P^{m+k}(x, A) = \int_{\mathcal{X}} P^k(y, A) P^m(x, dy),$$

donnant les équations de Kolmogorov $P^{m+k} = P^m P^k$

A.2 Propriétés

A.2.1 Homogénéité

Une chaîne de Markov (X_n) est dite **homogène** (requis par les algorithmes MCMC pour que s'appliquent les propriétés de convergence) si pour tout entier k , la loi conditionnelle de $(X_{n_1}, \dots, X_{n_k})$ sachant X_{n_0} est la même que celle de $(X_{n_1-n_0}, \dots, X_{n_k-n_0})$ sachant X_0 , avec $n_0 \leq n_1 \leq \dots \leq n_k$.

A.2.2 Irréductibilité

On appelle temps d'arrêt en $A \in \mathcal{B}$, l'entier

$$\tau_A = \inf\{n \geq 1; X_n = x_n \in A\},$$

qui par convention, est $+\infty$ si $x_n \notin A$, pour tout n . On note $L(x, A) = P_x(\tau_A < \infty)$, la probabilité de retour en A en un nombre fini d'étapes. Étant donné une mesure ϕ , une chaîne de Markov (X_n) est dite ϕ -**irréductible**, si pour tout $A \in \mathcal{B}, \phi(A) > 0$ entraîne

$$L(x, A) > 0, \quad \forall x \in \mathcal{X}.$$

A.2.3 Apériodicité

Un ensemble C est petit s'il existe $m \in \mathbf{N}^*$ et une mesure ν_m tels que

$$P^m(x, A) \geq \nu_m(A), \quad \forall x \in C, \forall A \in \mathcal{B}.$$

On dit qu'une chaîne de Markov (X_n) , ϕ -irréductible, admet un cycle de longueur d , s'il existe un ensemble petit C , d'entier et de loi de probabilité associés M et ν_M tels que d soit le p.g.d.c de

$$\{m \geq 1; \exists \delta_m > 0 \text{ tel que } C \text{ est petit pour } \nu_m \geq \delta_m \nu_M\}.$$

La chaîne est dite **apériodique** si $d = 1$.

A.2.4 Récurrence

Soient

$$\eta_A = \sum_{n=1}^{\infty} \mathbb{I}_A(X_n = x_n), \quad \text{et}$$

$$U(x, A) = \mathbb{E}_x[\eta_A]$$

respectivement le nombre de passage de (X_n) en A et le nombre moyen de passages en A . La chaîne (X_n) est **récurrente** s'il existe une mesure ϕ telle qu'elle soit ϕ -irréductible et si $U(x, A) = \infty$ pour tout $x \in \mathcal{X}$ et pour tout $A \in \mathcal{B}$ tel que $\phi(A) > 0$.

A.3 loi stationnaire

Une mesure σ -fine π est invariante pour le noyau de transition $P(., .)$ si

$$\pi(B) = \int_{\mathcal{X}} P(x, B) \pi(dx), \quad \forall B \in \mathcal{B}.$$

Lorsque π est une loi de probabilité, on parle de **loi stationnaire**. π est telle que si $X_0 \sim \pi$, alors $X_n \sim \pi$ pour tout n .

A.4 Ergodicité et convergence

L'existence d'une loi stationnaire conduit à considérer des conditions pour lesquelles (X_n) soit asymptotiquement distribuée selon cette loi telles que par exemple que la convergence se fasse indépendamment des conditions initiales, c'est à dire qu'à partir d'un certain rang n_0 (qu'on peut supposer égal 1), pour tout $n \geq n_0$, $X_n \sim \pi$, la loi stationnaire. On dit que la chaîne est **ergodique**. Et grâce aux théorèmes ergodiques, la quantité $\frac{1}{N} \sum_{n=n_0}^N h(x_n)$, quand $N \rightarrow +\infty$, tend presque-sûrement vers

$$\int_{\mathcal{X}} h(x) \pi(dx),$$

où h est une fonction bornée.

Quelques fonctions et distributions

Le but de ce chapitre est de présenter quelques fonctions et distributions de probabilité que nous avons utilisées ou avons fait cas dans notre travail de modélisation et éventuellement montrer les liens qui existent entre elles.

B.1 Fonction gamma (Γ) -Fonction de Bessel modifiée

B.1.1 Fonction Gamma (Γ)

La fonction Gamma sur l'ensemble des réels positifs non nuls est donnée par

$$\Gamma(r) = \int_0^{+\infty} t^{r-1} \exp(-t) dt$$

Elle est utilisée comme constante dans certaines densités de probabilité et possède entre autres comme propriétés :

$$\begin{aligned} \Gamma(r+1) &= r\Gamma(r) \quad \forall r > 0 \\ \Gamma(n+1) &= n! \quad \forall n \text{ entier naturel.} \end{aligned}$$

B.1.2 Fonction de Bessel modifiée

La fonction de Bessel modifiée est donnée par

$$K_\mu(y) = \frac{\sqrt{\pi}}{2^\nu \Gamma(\mu + \frac{1}{2})} y^\mu \int_0^{+\infty} (x^2 - 1)^{\mu - \frac{1}{2}} \exp(-yx) dx, \quad \mu > -\frac{1}{2}.$$

Elle est utilisée dans les fonctions de variance-covariance de Matèrn.

B.2 Quelques distributions

Il s'agit essentiellement des lois de probabilité que nous avons utilisées dans nos articles sans détails.

Loi uniforme

Une variable aléatoire réelle X suit une loi uniforme sur l'intervalle $[a,b]$ (a,b des réels), si sa mesure de probabilité admet pour densité :

$$f(x) = \frac{1}{b-a}$$

sur cet intervalle. Elle a pour moyenne $\frac{a+b}{2}$ et variance $\frac{(a-b)^2}{12}$. Elle est utilisée pour modéliser des phénomènes qu'on pense uniformément répartis. La loi uniforme est aussi utilisée en statistique bayésienne pour traduire un manque d'information *a priori*.

B.2.1 Loi gaussienne et loi log-normale

Une variable aléatoire réelle X suit une loi normale ou loi gaussienne, si sa mesure de probabilité admet pour densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad \forall x \in R,$$

avec m l'espérance mathématique de X et σ^2 , sa variance. Elle s'utilise pour la modélisation de beaucoup de phénomènes dont notamment celui dépendant d'un grand nombre de facteurs agissant sous forme additive, chaque facteur ayant une faible variance par rapport à la résultante. Elle est aussi forme limite de nombreuses distributions discrètes. Une loi dérivée de cette dernière, dite loi log-normale, est celle de la variable $Y = \exp(X)$. Elle admet pour densité :

$$g(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{\log(x-m)^2}{2\sigma^2}\right), \quad y \geq 0.$$

Elle est la loi d'un phénomène dépendant d'un grand nombre de facteurs agissant sous forme multiplicative.

On parlera de loi normale multidimensionnelle ou multivariée, si elle est la loi d'un vecteur aléatoire X , transposé de (X_1, \dots, X_n) . Dans ce cas, elle est de la forme :

$$f(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right),$$

où $x \in R^n$, $\mu \in R^n$, vecteur-centre de la distribution et Σ , matrice de variance-covariance, inversible et de déterminant $|\Sigma|$. Elle est utilisée pour beaucoup de processus aléatoires, de fois sans aucune justification physique ou réelle.

B.2.2 Loi Inverse-gamma

Une variable aléatoire réelle positive X suit une loi *gamma* de paramètres t et λ positifs, si sa mesure de probabilité admet pour densité

$$f(x; t, \lambda) = \frac{\lambda \exp(-\lambda x) (\lambda x)^{t-1}}{\Gamma(t)}.$$

La loi Inverse-Gamma est la loi de la variable aléatoire $Y = \frac{1}{X}$. Sa densité est :

$$g(y; t, \lambda) = \frac{\lambda^t}{\Gamma(t)} (1/y)^{t+1} \exp(-\lambda/y).$$

Elles sont par exemple utilisées en inférence bayésienne pour avoir des lois conjuguées.

B.2.3 Loi de Poisson- Loi Binomiale négative

La loi de Poisson est la loi des événements rares, c'est-à-dire loi des petits nombres. Une variable aléatoire discrète X suivant une loi de Poisson a pour distribution

$$P(X = k) = \frac{\lambda^k}{k!} \exp(-\lambda)$$

où k est un entier naturel et λ positif, paramètre de la loi, qui est à la fois sa moyenne et sa variance. Mais il arrive de fois que la moyenne des données qu'on veut modéliser avec une loi de Poisson soit différente de leur variance ou soit aléatoire. Alors on fait recours à un mélange de Poisson-gamma définie par :

$$P(Y = k) = \int_0^{+\infty} P(X = k) f(\lambda; r; \theta) d\lambda$$

où

$$P(X = k) = \frac{\lambda^k}{k!} \exp(-\lambda)$$

et

$$f(\lambda; r; \theta) = \frac{\lambda^{r-1} \exp(-\lambda/\theta)}{\Gamma(r) \theta^r}.$$

C'est une forme généralisée de la loi binomiale négative. En effet, en posant $t = \lambda \frac{\theta + 1}{\theta}$, on aura

$$\begin{aligned} P(Y = k) &= \left(\frac{\theta + 1}{\theta}\right)^{r+k} \cdot \frac{1}{\Gamma(r) k! \theta^r} \int_0^{+\infty} t^{k+r-1} \exp(-t) dt \\ &= \frac{\Gamma(k+r)}{\Gamma(r) k!} \left(\frac{1}{\theta + 1}\right)^r \left(\frac{\theta}{\theta + 1}\right)^k. \end{aligned}$$

Puis, en posant $p = \frac{1}{\theta + 1}$ et $q = \frac{\theta}{\theta + 1}$, on a

$$P(Y = k) = \frac{\Gamma(k + r)}{\Gamma(r)k!} p^r q^k$$

qui quand $r = n$, entier naturel, équivaut à la loi binomiale négative définie par

$$P(Y = k) = C_{n+k-1}^k p^n q^k, \quad \text{avec } q = 1 - p,$$

car pour tout entier naturel on a

$$\frac{(k + n - 1)!}{k!(n - 1)!} = \frac{\Gamma(k + n)}{\Gamma(n)k!}.$$

Ce mélange tend vers une loi de Poisson quand $r \rightarrow +\infty$. En effet, si cette fois, on pose $\lambda = r \frac{1-p}{p}$, c'est-à-dire $p = \frac{r}{\lambda + r}$, la loi binomiale négative se réécrit sous la forme

$$g_{\lambda,r}(k) = \frac{\lambda^k}{k!} \frac{A_{r+k-1}^k}{(r + \lambda)^k} \frac{1}{\left(1 + \frac{\lambda}{r}\right)^r}.$$

Et quand $r \rightarrow +\infty$, on a :

$$g_{\lambda,r}(k) = \frac{\lambda^k}{k!} \cdot 1 \cdot \frac{1}{\exp(\lambda)},$$

qui est une loi de Poisson de paramètre λ .

ANNEXE C

Données d'illustration

Ici nous présentons un exemple de données telles que nous les avons utilisées dans notre étude. Il s'agit d'admissions dus à l'asthme dans 4 hôpitaux de l'Hérault en 2009. Nous avons le nombre de patients par lieu résidence et par hôpital avec leur coordonnées géographiques (latitude, longitude)

Feuille1

Nomcommunes	Latitude	Longitude	CHUMontp	Millénaire	CHBéziers	CHBassin
Abeilhan	43,45	3,3	0	0	3	0
Adissan	43,54	3,433	1	0	1	0
Agde	43,314	3,471	2	0	8	10
Agel	43,338	2,852	0	0	0	0
Agonès	43,89	3,73	1	0	0	0
Aigne	43,333	2,796	0	0	0	0
Aigues.Vives	43,338	2,818	0	0	0	0
Les Aires	43,584	3,079	1	0	1	0
Alignan.du.Vent	43,469	3,341	0	0	0	0
Aniane	43,685	3,587	1	0	0	0
Arboras	43,711	3,844	0	0	0	0
Argelliers	43,697	3,672	1	0	0	0
Aspiran	43,567	3,45	2	0	1	0
Assas	43,702	3,9	2	0	0	0
Assignan	43,398	2,889	0	0	1	0
Aumelas	43,605	3,599	0	0	0	0
Aumes	43,467	3,463	0	0	0	0
Autignac	43,5	3,168	0	0	0	0
Avène	43,756	3,11	1	0	0	0
Azillanet	43,324	3,737	0	0	0	0
Babeau Bouldoux	43,436	2,908	0	0	0	0
Baillargues	43,66	4,014	0	0	0	0
Balaruc les Bains	43,442	3,68	0	1	0	6
Balaruc le Vieux	43,46	3,684	0	0	0	0
Bassan	43,412	3,254	0	0	0	0
Beaufort	43,298	2,757	0	0	0	0
Beaulieu	43,729	4,022	2	1	0	0
Bédarieux	43,616	3,158	0	0	0	0
Bélarga	43,552	3,488	0	0	0	0
Berlou	43,491	2,956	0	0	0	0
Bessan	43,631	3,424	0	0	0	0
Béziers	43,345	3,218	1	0	24	0
Boisseron	43,76	4,08	0	0	0	0
Boisset	43,404	2,701	0	0	0	0
La Boissière	43,664	3,645	0	0	0	0
Le Bosc	43,711	3,384	0	0	0	0
Boujan sur Libron	43,372	3,249	0	0	0	0
Le Bousquet d'Orb	43,694	3,164	0	0	0	0
Bouzigues	43,449	3,659	0	0	0	2
Brenas	43,655	3,257	0	0	0	0
Brignac	43,623	3,473	0	0	1	0
Brissac	43,878	3,702	0	0	0	0
Buzignargues	43,77	4,003	0	0	0	0
Cabrerolles	43,546	3,124	0	0	0	0
Cabrières	43,576	3,361	0	0	0	0
Cambon et Salvergues	43,617	2,859	0	0	1	0
Campagnan	43,539	3,498	0	0	0	0
Campagne	43,789	4,029	0	0	0	0
Camplong	43,673	3,117	0	0	0	0
Candillargues	43,619	4,071	1	1	0	0
Canet	43,602	3,491	0	0	0	0
Capestang	43,328	3,044	0	0	2	0

Feuille1

Carlenças et Levas	43,633	3,227	0	0	0	0
Cassagnoles	43,387	2,619	0	0	0	0
Castanet le Haut	43,689	2,954	0	0	0	0
Castelnau de Guers	43,436	3,439	0	0	2	0
Castelnau le Lez	43,633	3,897	1	1	0	0
Castries	43,677	3,986	0	0	0	0
La Caunette	43,353	2,778	0	0	0	0
Causse de la Selle	43,804	3,648	0	0	0	0
Causses et Veyran	43,475	3,086	0	0	3	0
Caussiniojols	43,552	3,151	0	0	0	0
Caux	43,507	3,368	0	0	1	0
Le Caylar	43,862	3,317	0	0	0	0
Cazedarnes	43,423	3,023	0	0	0	0
Cazevieille	43,769	3,787	3	0	0	0
Cazilhac	43,918	3,696	0	0	0	0
Cazouls dHérault	43,507	3,457	0	0	0	0
Cazouls lès Béziers	43,393	3,102	0	0	3	0
Cébazan	43,404	2,973	0	0	0	0
Ceilhes et Rocozels	43,804	3,111	0	0	0	0
Celles	43,663	3,34	0	0	0	0
Cers	43,323	3,306	0	0	2	0
Cessenon sur Orb	43,45	3,05	0	0	0	0
Cesseroas	43,326	2,714	0	0	0	0
Ceyras	43,646	3,457	0	0	0	0
Clapiers	43,657	3,889	2	0	0	0
Claret	43,861	3,905	0	0	0	0
Clermont IHérault	43,628	3,429	0	0	0	0
Colombières sur Orb	43,58	3,004	0	0	0	0
Colombiers	43,313	3,14	0	0	1	0
Combailiaux	43,678	3,767	1	0	0	0
Combes	43,604	3,046	0	0	4	0
Corneilhan	43,401	3,192	0	0	0	0
Coulobres	43,451	3,278	0	0	0	0
Courniou	43,474	2,711	0	0	0	0
Cournonsec	43,548	3,704	1	0	0	0
Cournonterral	43,558	3,721	0	0	0	0
Creissan	43,374	3,01	0	0	0	0
Le Crès	43,651	3,937	0	0	0	0
Le Cros	43,868	3,364	0	0	0	0
Cruzy	43,356	2,941	0	0	0	0
Dio et Valquières	43,669	3,209	0	0	0	0
Espondeilhan	43,44	3,262	0	0	0	0
Fabrègues	43,551	3,777	2	2	0	0
Faugères	43,566	3,188	0	0	0	0
Félines Minervoias	43,33	2,6	0	0	0	0
Ferrals les Montagnes	43,404	2,631	0	0	0	0
Ferrières les Verreries	43,877	3,798	0	0	0	0
Ferrières Poussarou	43,486	2,899	0	0	0	0
Florensac	43,383	3,463	1	0	3	0
Fontanès	43,795	3,914	0	0	0	0
Fontès	43,54	3,38	0	0	1	0
Fos	43,567	3,245	0	0	0	0
Fouzilhon	43,502	3,246	0	0	0	0

Feuille1

Fozières	43,752	3,357	0	0	0	0
Fraisse sur Agout	43,607	2,797	0	0	0	0
Frontignan	43,447	3,756	2	1	0	12
Gabian	43,513	3,273	0	0	0	0
Galargues	43,77	4,02	0	0	0	0
Ganges	43,935	3,704	0	0	0	0
Garrigues	43,794	4,013	0	0	0	0
Gigean	43,501	3,712	0	0	0	0
Gignac	43,653	3,552	0	0	0	0
Gorniès	43,889	3,622	0	0	0	0
Grabels	43,649	3,797	0	0	0	0
Graissessac	43,682	3,094	0	0	0	0
Guzargues	43,721	3,926	0	0	0	0
Hérépian	43,594	3,115	0	0	0	0
Jacou	43,661	3,912	0	0	0	0
Joncels	43,737	3,193	0	0	0	0
Jonquières	43,676	3,476	0	0	0	0
Juvignac	43,613	3,818	3	0	0	0
Lacoste	43,647	3,436	0	0	0	0
Lagamas	43,674	3,523	0	0	0	0
Lamalou les Bains	43,596	3,081	0	0	0	0
Lansargues	43,651	4,074	0	0	0	0
Laroque	43,923	3,728	0	0	0	0
Lattes	43,568	3,901	3	2	0	0
Laurens	43,524	3,197	0	0	0	0
Lauret	43,836	3,884	0	0	0	0
Lauroux	43,778	3,289	0	0	0	0
Lavalette	43,692	3,271	0	0	0	0
Lavérune	43,585	3,803	0	1	0	0
Lespignan	43,274	3,174	0	0	3	0
Lézignan la Cèbe	43,494	3,439	0	0	0	0
Liausson	43,636	3,369	0	0	0	0
Lieuran Cabrières	43,585	3,415	0	0	0	0
Lieuran lès Béziers	43,419	3,237	0	0	0	0
Lignan sur Orb	43,389	3,168	0	0	0	0
La Livinière	43,318	3,636	0	0	0	0
Lodève	43,734	3,323	0	0	0	0
Loupian	43,449	3,615	0	0	0	0
Lunas	43,707	3,196	0	0	0	0
Lunel	43,676	4,134	6	0	0	0
Lunel Viel	43,679	4,092	0	0	0	0
Magalas	43,472	3,222	0	0	0	0
Maraussan	43,369	3,158	0	0	0	0
Margon	43,487	3,307	0	0	0	0
Marseillan	43,357	3,528	0	0	1	0
Marsillargues	43,665	4,179	1	0	0	0
Mas de Londres	43,781	3,752	0	0	0	0
Les Matelles	43,731	3,809	0	0	0	0
Mauguio	43,615	4,01	0	0	0	0
Maureilhan	43,358	3,12	0	0	0	0
Mérifons	43,633	3,295	0	0	0	0
Mèze	43,428	3,605	0	0	0	0
Minerve	43,355	2,744	0	0	0	0

Feuille1

Mireval	43,508	3,801	0	0	0	0
Mons	43,572	2,955	0	0	0	0
Montady	43,331	3,12	0	0	0	0
Montagnac	43,481	3,485	0	0	0	0
Montarnaud	43,649	3,698	5	1	0	0
Montaud	43,753	3,955	0	0	0	0
Montbazin	43,514	3,696	1	0	0	3
Montblanc	43,397	3,368	0	0	0	0
Montels	43,294	3,032	0	0	0	0
Montesquieu	43,557	3,255	0	0	0	0
Montferrier sur Lez	43,667	3,859	0	0	0	0
Montouliers	43,338	2,908	0	0	0	0
Montoulieu	43,928	3,792	0	0	0	0
Montpellier	43,6	3,883	46	7	0	0
Montpeyroux	43,695	3,506	0	0	0	0
Moulès et Baucels	43,948	3,731	0	0	0	0
Mourèze	43,619	3,356	0	0	0	0
Mudaison	43,648	4,041	0	0	0	0
Murles	43,688	3,747	0	0	0	0
Murviel lès Béziers	43,441	3,144	0	0	0	0
Murviel lès Montpellier	43,604	3,736	0	0	0	0
Nébian	43,607	3,432	0	0	0	0
Neffiès	43,534	3,33	0	0	0	0
Nézignan l'Évêque	43,421	3,405	0	0	0	0
Nissan lez Enserune	43,29	3,129	0	0	0	0
Nizas	43,513	3,408	0	0	0	0
Notre Dame de Londres	43,827	3,777	0	0	0	0
Octon	43,655	3,304	0	0	0	0
Olargues	43,557	2,913	0	0	0	0
Olmet et Villecun	43,708	3,287	0	0	0	0
Olonzac	43,284	2,731	0	0	0	0
Oupia	43,291	2,766	0	0	0	0
Pailhès	43,431	3,187	0	0	0	0
Palavas les Flots	43,529	3,936	0	1	0	0
Pardailhan	43,451	2,846	0	0	0	0
Paulhan	43,54	3,458	0	0	0	0
Pégairolles de Buèges	43,806	3,586	0	0	0	0
Pégairolles de l'Escalette	43,803	3,323	0	0	0	0
Péret	43,576	3,394	0	0	0	0
Pérols	43,562	3,954	2	0	0	0
Pézenas	43,46	3,425	0	0	0	0
Pézènes les Mines	43,59	3,252	0	0	0	0
Pierrerue	43,428	2,978	0	0	0	0
Pignan	43,583	3,765	0	0	0	0
Pinet	43,405	3,56	0	0	0	0
Plaisan	43,56	3,527	0	0	0	0
Les Plans	43,755	3,269	0	0	0	0
Poilhes	43,309	3,075	0	0	0	0
Pomérols	43,39	3,499	0	0	0	0
Popian	43,62	3,535	0	0	0	0
Portiragnes	43,304	3,336	0	0	0	0
Le Pouget	43,593	3,524	0	0	0	0
Le Poujol sur Orb	43,579	3,06	0	0	0	0

Feuille1

Poujols	43,765	3,324	0	0	0	0
Poussan	43,489	3,671	0	0	0	0
Pouzolles	43,484	3,28	0	0	0	0
Pouzols	43,618	3,513	0	0	0	0
Le Pradal	43,628	3,103	0	0	0	0
Prades le Lez	43,698	3,863	1	0	0	0
Prades sur Vernazobre	43,447	2,986	0	0	0	0
Prémian	43,53	2,832	0	0	0	0
Le Puech	43,695	3,313	0	0	0	0
Puéchabon	43,715	3,619	0	0	0	0
Puilacher	43,568	3,507	0	0	0	0
Puimisson	43,44	3,207	0	0	0	0
Puissalicon	43,459	3,236	0	0	0	0
Puisserguier	43,369	3,039	0	0	1	0
Quarante	43,347	2,962	0	0	0	0
Restinclières	43,723	4,037	0	0	0	0
Rieussec	43,418	2,739	0	0	0	0
Riols	43,506	2,79	0	0	0	0
Les Rives	43,843	3,267	0	0	0	0
Romiguières	43,818	3,229	0	0	0	0
Roquebrun	43,502	3,029	0	0	0	0
Roqueredonde	43,801	3,21	0	0	0	0
Roquessels	43,552	3,224	0	0	0	0
Rosis	43,623	3,005	0	0	0	0
Rouet	43,818	3,792	0	0	0	0
Roujan	43,505	3,311	0	0	0	0
Saint André de Buèges	43,849	3,663	0	0	0	0
Saint André de Sangonis	43,65	3,503	0	0	0	0
Saint Aunès	43,638	3,966	0	0	0	0
Saint Bauzille de la Sylve	43,618	3,546	0	0	0	0
Saint Bauzille de Montmel	43,769	3,955	0	0	0	0
Saint Bauzille de Putois	43,895	3,734	0	0	0	0
Saint Brès	43,666	4,03	0	0	0	0
Saint Chinian	43,422	2,946	0	0	0	0
Saint Christol	43,727	4,079	0	0	0	0
Saint Clément de Rivière	43,686	3,851	0	0	0	0
Sainte Croix de Quintillargue	43,773	3,909	0	0	0	0
Saint Drézéry	43,734	3,984	0	0	0	0
Saint Étienne d Albagnan	43,532	2,854	0	0	0	0
Saint Étienne de Gourgas	43,771	3,377	0	0	0	0
Saint Étienne Estréchoux	43,665	3,107	0	0	0	0
Saint Félix de IHéras	43,834	3,3002	0	0	0	0
Saint Félix de Lodez	43,662	3,459	0	0	0	0
Saint Gély du Fesc	43,693	3,806	0	0	0	0
Saint Geniès des Mourgues	43,697	4,037	0	0	0	0
Saint Geniès de Varensal	43,681	3,003	0	0	0	0
Saint Geniès de Fontedit	43,468	3,179	0	0	0	0
Saint Georges dOrques	43,611	3,783	1	0	0	0
Saint Gervais sur Mare	43,653	3,041	0	0	0	0
Saint Guilhem le Désert	43,734	3,549	0	0	0	0
Saint Guiraud	43,677	3,454	0	0	0	0
Saint Hilaire de Beauvoir	43,757	4,012	0	0	0	0
Saint Jean de Buèges	43,829	3,618	0	0	0	0

Feuille1

Saint Jean de Cornies	43,744	4,002	0	0	0	0
Saint Jean de Cuculles	43,753	3,836	0	0	0	0
Saint.Jean de Fos	43,702	3,551	0	0	0	0
Saint Jean de la Blaquièrre	43,716	3,423	0	0	0	0
Saint Jean de Minervoies	43,389	2,831	0	0	0	0
Saint Jean de Védas	43,557	3,826	2	0	0	0
Saint Julien	43,571	2,926	0	0	0	0
Saint Just	43,657	4,115	0	0	0	0
Saint Martin de l'Arçon	43,576	2,983	0	0	0	0
Saint Martin de Londres	43,791	3,734	0	0	0	0
Saint Mathieu de Trévièrs	43,769	3,858	0	0	0	0
Saint Maurice Navacelles	43,846	3,518	0	0	0	0
Saint Michel	43,855	3,387	0	0	0	0
Saint Nazaire de Ladarez	43,509	3,075	0	0	0	0
Saint Nazaire de Pézan	43,644	4,119	0	0	0	0
Saint Pargoire	43,528	3,519	0	0	0	0
Saint Paul et Valmalle	43,626	3,675	0	0	0	0
Saint Pierre de la Fage	43,794	3,426	0	0	0	0
Saint Pons de Thomières	43,489	2,762	0	0	0	0
Saint Pons de Mauchiens	43,513	3,514	0	0	0	0
Saint Privat	43,754	3,424	0	0	0	0
Saint Saturnin de Lucian	43,695	3,472	0	0	0	0
Saint Sériès	43,732	4,106	0	0	0	0
Saint Thibéry	43,398	3,418	0	0	1	0
Saint Vincent de Barbeyrarg	43,706	3,878	0	0	0	0
Saint Vincent d Olargues	43,562	2,879	0	0	0	0
Salasc	43,621	3,316	0	0	0	0
La Salvetat sur Agout	43,602	2,703	0	0	0	0
Saturargues	43,723	4,113	0	0	0	0
Saussan	43,573	3,774	0	0	0	0
Saussines	43,764	4,057	0	0	0	0
Sauteyrargues	43,838	3,92	0	0	0	0
Sauvian	43,295	3,263	0	0	3	0
Sérignan	43,283	3,281	0	0	0	0
Servian	43,427	3,303	0	0	0	0
Sète	43,407	3,696	3	0	0	38
Siran	43,314	2,66	0	0	0	0
Sorbs	43,892	3,399	0	0	0	0
Soubès	43,768	3,345	0	0	0	0
Le Soulié	43,551	2,691	0	0	0	0
Soumont	43,728	3,353	0	0	0	0
Sussargues	43,712	4,003	0	0	0	0
Taussac la Billière	43,622	3,072	0	0	0	0
Teyran	43,684	3,927	0	0	0	0
Thézan lès Béziers	43,422	3,17	0	0	0	0
Tourbes	43,447	3,379	0	0	0	0
La Tour sur Orb	43,653	3,149	0	0	0	0
Tressan	43,575	3,491	0	0	0	0
Le Triadou	43,739	3,853	0	0	0	0
Uscas d'Hérault	43,519	3,464	0	0	0	0
Uscas du Bosc	43,724	3,403	0	0	0	0
La Vacquerie et Saint Martir	43,791	3,46	1	0	0	0
Vacquières	43,844	3,945	0	0	0	0

Feuille1

Vailhan	43,553	3,303	0	0	0	0
Vailhauquès	43,672	3,718	0	0	0	0
Valergues	43,667	4,062	0	0	0	0
Valflaunès	43,801	3,872	0	0	0	0
Valmascle	43,596	3,298	0	0	0	0
Valras.Plage	43,247	3,292	0	0	3	0
Valros	43,42	3,367	0	0	0	0
Vélieux	43,388	2,731	0	0	0	0
Vendargues	43,657	3,969	1	0	0	0
Vendémian	43,581	3,562	0	0	0	0
Vendres	43,27	3,226	1	0	0	0
Vérargues	43,716	4,1	0	0	0	0
Verreries de Moussans	43,45	2,686	0	0	0	0
Vias	43,314	3,418	0	0	1	0
Vic la Gardiole	43,491	3,797	0	0	0	0
Vieussan	43,541	2,977	0	0	0	0
Villemagne l'Argentière	43,619	3,118	0	0	0	0
Villeneuve lès Béziers	43,315	3,281	0	0	0	0
Villeneuve lès Maguelone	43,532	3,862	2	0	0	0
Villeneuvevette	43,611	3,4	0	0	0	0
Villespassans	43,381	2,916	0	0	0	0
Villetelle	43,734	4,14	0	0	0	0
Villeveyrac	43,501	3,607	0	0	0	0
Viols en Laval	43,754	3,725	0	0	0	0
Viols le Fort	43,744	3,704	0	0	0	0
La Grande Motte	43,558	4,089	2	3	0	0

Bibliographie

- [Agarwal et al., 2002] Agarwal, D. K., Gelfand, A. E., and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological statistics*, 9(4) :341–355.
- [Aho, 2015] Aho, K. (2015). *asbio : A Collection of Statistical Tools for Biologists*. R package version 1.1-5.
- [Aubas et al., 2013] Aubas, C., Bourdin, A., Aubas, P., Gamez, A., Halimi, L., Vachier, I., Malafaye, N., Chanez, P., and Molinari, N. (2013). Role of comorbid conditions in asthma hospitalizations in the south of france. *Allergy*, 68(5) :637–643.
- [Banerjee et al., 2004] Banerjee, S., Carlin, B., and Gelfand, A. (2004). Hierarchical modeling and analysis for spatial statistics. *New York : Chapman&Hall/CRC*.
- [Bayes, 1763] Bayes, T. (1763). A letter from the late reverend mr. thomas bayes, frs to john canton, ma and frs. *Philosophical Transactions (1683-1775)*, 53 :269–271.
- [Ben-Akiva, 1972] Ben-Akiva, M. (1972). The structure of travel demand models, transportation systems division, department of civil engineering.
- [Berger et al., 2001] Berger, J. O., De Oliveira, V., and Sansó, B. (2001). Objective bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456) :1361–1374.
- [Bounnafous and Saccareau, 2012] Bounnafous, F. and Saccareau, M. (2012). Analyse de l’attractivité des hôpitaux publics et privés sur les malades par methode mcmc. Rapport de stage Master 1 inédit, UM2, Montpellier, France.
- [Box and Cox, 1964] Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252.
- [Boyd and Mellman, 1980] Boyd, J. H. and Mellman, R. E. (1980). The effect of fuel economy standards on the us automotive market : an hedonic demand analysis. *Transportation Research Part A : General*, 14(5-6) :367–378.

- [Cardell and Dunbar, 1980] Cardell, N. S. and Dunbar, F. C. (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research Part A : General*, 14(5-6) :423–434.
- [Carlin and Louis, 2000] Carlin, B. P. and Louis, T. A. (2000). Bayes and empirical bayes methods for data analysis.
- [Chevreul et al., 2010] Chevreul, K., Durand-Zaleski, I., Bahrami, S., Hernández-Quevedo, C., and Mladovsky, P. (2010). Health system review. *Erasmus*, 12 :93–103.
- [Cohen, 1963] Cohen, A. C. (1963). *Estimation in mixtures of discrete distributions*. Statistical Pub. Society.
- [Cressie, 1993] Cressie, N. (1993). Statistics for spatial data : Wiley series in probability and statistics. *Wiley-Interscience, New York*, 15 :105–209.
- [Cressie, 2015] Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.
- [de PALMA and Thisse, 1989] de PALMA, A. and Thisse, J.-F. (1989). Les modèles de choix discrets. *Annales d’Economie et de Statistique*, pages 151–190.
- [Debreu, 1960] Debreu, G. (1960). Individual choice behavior—a theoretical-analysis-luce, rd.
- [Gelfand and Smith, 1990] Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410) :398–409.
- [Gelman et al., 2004] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). Bayesian data analysis. texts in statistical science series.
- [Gneiting, 1997] Gneiting, T. (1997). *Symmetric positive definite functions with applications in spatial statistics*.
- [Greene, 1994] Greene, W. H. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models.
- [Johnson and Kotz, 1969] Johnson, N. L. and Kotz, S. (1969). *Distributions in statistics*. J. Wiley.
- [Jung et al., 2011] Jung, K., Feldman, R., and Scanlon, D. (2011). Where would you go for your next hospitalization? *Journal of health economics*, 30(4) :832–841.
- [Kahneman and Tversky, 1972] Kahneman, D. and Tversky, A. (1972). Subjective probability : A judgment of representativeness. *Cognitive psychology*, 3(3) :430–454.
- [Kass et al., 1999] Kass, D., Thurston, H., Schlather, K., Buck, L., Lassoie, J., Fernandes, E., et al. (1999). Sustainable mulch-based cropping systems with trees. *Agroforestry in sustainable agricultural systems*, pages 361–379.
- [Lako and Rosenau, 2009] Lako, C. J. and Rosenau, P. (2009). Demand-driven care and hospital choice. dutch health policy toward demand-driven care : results from a survey into hospital choice. *Health Care Analysis*, 17(1) :20–35.

- [Lambert, 1992] Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1) :1–14.
- [Laverty et al., 2012] Laverty, A. A., Smith, P. C., Pape, U. J., Mears, A., Wachter, R. M., and Millett, C. (2012). High-profile investigations into hospital safety problems in england did not prompt patients to switch providers. *Health Affairs*, 31(3) :593–601.
- [Luce, 1959] Luce, R. D. (1959). On the possible psychophysical laws. *Psychological review*, 66(2) :81.
- [Luce and Suppes, 1965] Luce, R. D. and Suppes, P. (1965). *Preference, utility, and subjective probability*. Wiley.
- [Marschak, 1960] Marschak, J. (1960). Stanford symposium on mathematical methods in the social sciences (ed.) k arrow.
- [Marschak et al., 1959] Marschak, J. et al. (1959). Binary choice constraints on random utility indicators. Technical report, Cowles Foundation for Research in Economics, Yale University.
- [Matérn, 1960] Matérn, B. (1960). *Spatial Variation : Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Statens skogsforskningsinstitut.
- [McFadden, 1974] McFadden, D. (1974). Vconditional logit analysis of qualitative choice behavior, *vinfrontiersineconometrics*, ed.byp.zarembka.
- [McFadden et al., 2000] McFadden, D., Train, K., et al. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5) :447–470.
- [Meyn and Tweedie, 1993] Meyn, S. P. and Tweedie, R. L. (1993). Markov chains and stochastic stability. communication and control engineering series. *Springer-Verlag London Ltd., London*, 1 :993.
- [Mullahy, 1986] Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, 33(3) :341–365.
- [Porell and Adams, 1995] Porell, F. W. and Adams, E. K. (1995). Hospital choice models : a review and assessment of their utility for policy impact analysis. *Medical Care Research and Review*, 52(2) :158–195.
- [Propper and Leckie, 2011] Propper, C. and Leckie, G. (2011). Increasing competition between providers in health care markets : the economic evidence.
- [Quandt, 1956] Quandt, R. E. (1956). A probabilistic theory of consumer behavior. *The Quarterly Journal of Economics*, pages 507–536.
- [R Core Team, 2017] R Core Team (2017). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [Robert, 1996] Robert, C. P. (1996). *Méthodes de Monte Carlo par chaînes de Markov*. Economica.

- [Schlather, 1999] Schlather, M. (1999). An introduction to positive definite functions and to unconditional simulation of random fields. Technical report, Technical report st 99-10, Dept. of Mathematics and Statistics, Lancaster University.
- [Schoenberg, 1938] Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3) :522–536.
- [Spiegelhalter et al., 2002] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 64(4) :583–639.
- [Thurstone, 1927] Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4) :273.
- [Train, 1986] Train, K. (1986). Qualitative choice analysis.
- [Train, 2009] Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- [Tversky, 1972] Tversky, A. (1972). Elimination by aspects : A theory of choice. *Psychological review*, 79(4) :281.
- [Tversky and Kahneman, 1973] Tversky, A. and Kahneman, D. (1973). Availability : A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2) :207–232.
- [Varkevisser et al., 2010] Varkevisser, M., van der Geest, S. A., and Schut, F. T. (2010). Assessing hospital competition when prices don’t matter to patients : the use of time-elasticities. *International Journal of Health Care Finance and Economics*, 10(1) :43–60.
- [Victoor et al., 2012] Victoor, A., Delnoij, D. M., Friele, R. D., and Rademakers, J. J. (2012). Determinants of patient choice of healthcare providers : a scoping review. *BMC health services research*, 12(1) :272.
- [Waller and Gotway, 2004] Waller, L. A. and Gotway, C. A. (2004). *Applied spatial statistics for public health data*, volume 368. John Wiley & Sons.
- [Zhang, 2004] Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465) :250–261.