

UNIVERSITÉ DE MONTRÉAL

**ANALYSE DU COMPORTEMENT HÉTÉROGÈNE DES
USAGERS DANS UN RÉSEAU**

PAR

Zacharie-Francis KLOK

DÉPARTEMENT D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE

FACULTÉ DES ARTS ET SCIENCES

MÉMOIRE DÉPOSÉ À LA FACULTÉ DES ARTS ET SCIENCES

EN VUE DE L'OBTENTION DU GRADE DE MAÎTRE ÈS. SCIENCES

EN INFORMATIQUE

OPTION RECHERCHE OPÉRATIONNELLE

AOÛT, 2014

©Zacharie-Francis KLOK, 2014

Résumé

Le nombre important de véhicules sur le réseau routier peut entraîner des problèmes d'encombrement et de sécurité. Les usagers des réseaux routiers qui nous intéressent sont les camionneurs qui transportent des marchandises, pouvant rouler avec des véhicules non conformes ou emprunter des routes interdites pour gagner du temps. Le transport de matières dangereuses est réglementé et certains lieux, surtout les ponts et les tunnels, leur sont interdits d'accès. Pour aider à faire appliquer les lois en vigueur, il existe un système de contrôles routiers composé de structures fixes et de patrouilles mobiles. Le déploiement stratégique de ces ressources de contrôle mise sur la connaissance du comportement des camionneurs que nous allons étudier à travers l'analyse de leurs choix de routes.

Un problème de choix de routes peut se modéliser en utilisant la théorie des choix discrets, elle-même fondée sur la théorie de l'utilité aléatoire. Traiter ce type de problème avec cette théorie est complexe. Les modèles que nous utiliserons sont tels, que nous serons amenés à faire face à des problèmes de corrélation, puisque plusieurs routes partagent probablement des arcs. De plus, puisque nous travaillons sur le réseau routier du Québec, le choix de routes peut se faire parmi un ensemble de routes dont le nombre est potentiellement infini si on considère celles ayant des boucles. Enfin, l'étude des choix faits par un humain n'est pas triviale. Avec l'aide du modèle de choix de routes retenu, nous pourrions calculer une expression de la probabilité qu'une route soit prise par le camionneur.

Nous avons abordé cette étude du comportement en commençant par un travail de description des données collectées. Le questionnaire utilisé par les contrôleurs permet de collecter des données concernant les camionneurs, leurs véhicules et le lieu du contrôle. La description des données observées est une étape essentielle, car elle permet de présenter clairement à un analyste potentiel ce qui est accessible pour étudier les comportements des camionneurs. Les données observées lors d'un contrôle constitueront ce que nous appellerons une observation. Avec les attributs du réseau, il sera possible de modéliser le réseau routier du Québec. Une sélection de certains attributs permettra de spécifier la fonction d'utilité et par conséquent la fonction permettant de calculer les probabilités de choix de routes par un camionneur. Il devient alors possible d'étudier un comportement en se basant sur des observations. Celles provenant du terrain ne nous donnent pas suffisamment d'information actuellement et même en spécifiant bien un modèle, l'estimation des paramètres n'est pas possible. Cette

dernière est basée sur la méthode du maximum de vraisemblance. Nous avons l'outil, mais il nous manque la matière première que sont les observations, pour continuer l'étude.

L'idée est de poursuivre avec des observations de synthèse. Nous ferons des estimations avec des observations complètes puis, pour se rapprocher des conditions réelles, nous continuerons avec des observations partielles. Ceci constitue d'ailleurs un défi majeur. Nous proposons pour ces dernières, de nous servir des résultats des travaux de (Bierlaire et Frejinger, 2008) en les combinant avec ceux de (Fosgerau, Frejinger et Karlström, 2013). Bien qu'elles soient de nature synthétiques, les observations que nous utilisons nous mèneront à des résultats tels, que nous serons en mesure de fournir une proposition concrète qui pourrait aider à optimiser les décisions des responsables des contrôles routiers. En effet, nous avons réussi à estimer, sur le réseau réel du Québec, avec un seuil de signification de 0,05 les valeurs des paramètres d'un modèle de choix de routes discrets, même lorsque les observations sont partielles. Ces résultats donneront lieu à des recommandations sur les changements à faire dans le questionnaire permettant de collecter des données.

Mots clés : logit récursif, choix discrets, choix de routes, utilité aléatoire, maximum de vraisemblance, optimisation non linéaire sans contraintes, observations partielles, étude du comportement.

Abstract

Using transportation roads enables workers to reach their work facilities. Security and traffic jam issues are all the more important given that the number of vehicles is always increasing and we will focus on merchandise transporters in this study. Dangerous items transportation is under strict control as it is for example forbidden for them to be carried through a tunnel or across a bridge. Some transporters may drive a vehicle that has defects or/and they may be taking some forbidden roads so as to reach their destination faster. Transportation of goods is regulated by the law and there exists a control system, whose purpose is to detect frauds and to make sure controlled vehicles are in order. The strategic deployment of control resources can be based on the knowledge of transporters behaviour, which is going to be studied through their route choice analysis. The number of routes can be unbounded especially if we consider loops, which leads to a complex problem to be solved. We can also mention issues closely related to route choice problem using discrete choice models such as correlation between routes sharing links and point out the fact that human decision process is not considered something easy.

A route choice problem can be modelled based on the random utility theory and as a consequence we will focus on the discrete choice models. We are going to use such model on the real road network of Quebec and we will derive an expression of the probability, for a transporter, to pick one route. We are going to explain the way we did our study. It started first by doing a data description job as we are convinced this is a step that will help other analysts to have a clear view of the data situation. Some data are network related and the corresponding attributes collected will be used to model the road network of Quebec. We will use some attributes to explain the utility function, which leads to the definition of the function that gives the probability that a user takes a given route. Once this function is fully specified, the behaviour study can be done, except that we have a set of observations that are absolutely incomplete. When observations are a gathering of data collected during a road control, the information they provide us is not enough and thus, the parameters estimation will fail.

We might seem blocked but in fact, we brought the idea of using simulated observations. We are going to estimate model parameters with firstly complete observations and in order to imitate the real conditions, we then are going to use partial observations. This constitutes a main challenge and we overcome it by using the results presented in ([Bierlaire et Frejinger, 2008](#)) combined with those

from (Fosgerau, Frejinger et Karlström, 2013). We will demonstrate that even though the observations used are simulated, we will deliver conclusions that can be useful for road network managers. The main results we provide in this work is that estimation can be done with a 0,05 signification level on real road network of Quebec, while the observations are incomplete. Eventually, our results should motivate network managers to improve the set of questions they use to collect data as it would help them to strengthen their knowledge about the merchandise transporters and hopefully, the decision process will lead to optimized resource deployments.

Mots clés: recursive logit, discrete choice, route choice, random utility, maximum likelihood, unconstrained and convex optimization, partial observation, behaviour analysis.

Sommaire

Chapter 1	Introduction	1
Chapter 2	Revue de littérature	4
2.1	Modèles de choix discrets	5
2.1.1	Théorie de l'utilité aléatoire	6
2.1.2	Dérivation de la probabilité de choix d'une alternative par un décideur	7
2.2	La propriété IIA du logit	13
2.3	Approximations	15
2.3.1	Homoscédasticité	15
2.3.2	Cas limites	15
2.3.3	Une fonction d'utilité linéaire par rapport aux paramètres	17
2.4	Estimation d'un modèle logit	17
2.4.1	Le maximum de vraisemblance	17
2.4.2	Résolution d'un problème d'optimisation non linéaire sans contraintes	20
2.4.3	Validation de l'estimateur du maximum de vraisemblance	22
2.5	Des défis inhérents aux problèmes de choix de routes	24
2.5.1	Des problèmes combinatoires et de corrélation	24
2.5.2	Les défis sur les observations	26
2.5.3	Techniques de collecte des observations	26
2.5.4	Traitement des observations partielles	28
2.6	Modèle logit récursif	29
2.6.1	Un choix de routes dynamique	29
2.6.2	Un modèle de choix de routes sans échantillonnages sur l'ensemble des alternatives	30
2.6.3	Prédiction de flux dans un réseau	33

2.6.4	Un correctif pour la corrélation	33
Chapter 3	Spécification du modèle	35
3.1	Récolte de données et notion d'observation	36
3.1.1	La récolte et la description de données	36
3.1.2	Les observations	38
3.2	Modèle	39
3.2.1	Modélisations du réseau et du comportement	39
3.2.2	Spécification de la fonction d'utilité déterministe	41
3.2.3	Spécification du terme d'erreur	43
Chapter 4	Résultats expérimentaux	45
4.1	Validation avec des observations synthétiques	45
4.1.1	Le sous-réseau	47
4.1.2	Génération des observations synthétiques complètes	49
4.1.3	Expression de la log-vraisemblance	49
4.1.4	Validation de l'estimateur du maximum de vraisemblance	50
4.2	Observations synthétiques partielles	52
4.2.1	Génération des observations synthétiques partielles	52
4.2.2	Fonction de log-vraisemblance modifiée	53
4.2.3	Validation de l'estimateur du maximum de vraisemblance	54
Chapter 5	Discussion	58

Liste des figures

2.1	La loi de Gumbel comme approximation de la loi normale centrée réduite	11
2.2	Cas limites	16
2.3	Test de Student bilatéral à un seuil de 0,05 de signification	24
2.4	Illustration de la notation	30
4.1	Sous-réseau sur l'A20	48
4.2	Sous-réseau sur l'A20	48
4.3	Observations partielles issues d'observations complètes	56

Liste des tableaux

I	Attributs du réseaux	38
II	Attributs du réseaux	39
III	Attributs provenant d'un MTV	40
IV	Validation de l'estimation des paramètres avec des observations complètes	51
V	Validation de l'estimation des paramètres avec des observations partielles	57

Nomenclature

Acronymes

BFGS Broyden-Fletcher-Goldfarb-Shanno

CIRRELT Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

CRQ Contrôle routier Québec

DPD Domaine de pertinence des données

EMV Estimateur du maximum de vraisemblance

GPS Global positioning system

IIA Independence of irrelevant alternatives

iid Indépendamment et identiquement distribuée

LS Link size

MTQ Ministère des transports du Québec

MTV Mouvement de transport vérifié

PSL Path size logit

Autres

∇ Opérateur de gradient

B_t Arc hessien calculé à l'itération t

$E[X]$ Espérance d'une variable aléatoire X

H_t La matrice hessienne à l'itération t

$L(\cdot)$ Fonction de vraisemblance

$LL(\cdot)$ Fonction de log-vraisemblance

Remerciements

Lorsqu'on arrive à la fin d'une maîtrise qui a demandé un important investissement personnel, on est animé par plusieurs sentiments qui vont du soulagement à la fierté en passant par la gratitude. Je l'exprime ici en remerciant en premier lieu Emma Frejinger, titulaire de la Chaire CN en intermodalité des transports, pour avoir dirigé mon travail de maîtrise et fait preuve de patience tout en ne cessant jamais de m'encourager. Il m'est arrivé souvent de perdre la globalité du projet en me concentrant trop sur des détails ou de partir dans de mauvaises directions et de manière subtile Emma me remettait sur la bonne voie en posant les bonnes questions. Je suis redevable envers elle pour m'avoir appris tant sur la modélisation du comportement avec la théorie des choix discrets et de m'avoir fait progresser tout au long de notre collaboration. Tack så mycket!

Je remercie Bernard Gendron, directeur du centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport (CIRRELT) pour m'avoir donné l'opportunité de travailler sur ce projet de maîtrise et de me rendre compte des interactions entre les chercheurs et les commanditaires de projets. C'est sa présentation du laboratoire lors du premier jour qui a orienté mon choix vers la recherche opérationnelle.

Le support de Lilianne Dupuis du CIRRELT et les discussions avec le sergent Krause de CRQ ont contribué à l'avancement du projet et je les remercie pour ça.

Je tiens à remercier Tien Anh Mai, qui m'a donné de bons conseils et le fait que j'ai pu utiliser son code m'a fait gagner beaucoup de temps.

Je remercie le corps administratif du CIRRELT qui a toujours répondu à mes demandes avec gentillesse et efficacité. Je remercie les contributeurs de l'application web ©OpenStreetMap pour l'utilisation sous licence de leurs cartes.

Il y a deux ans, j'ai pris la décision reprendre des études à l'Université de Montréal. J'ai tout de suite choisi de m'orienter vers la recherche opérationnelle car j'ai été séduit par le nombre et la nature des problèmes qu'on y aborde. Je voulais tester le système universitaire du Québec et j'en suis très satisfait. Je garde en souvenir les discussions passionnantes avec des gens brillants que j'ai rencontrés au CIRRELT, à l'Université de Montréal et en dehors, faisant de mon séjour au Québec un moment particulièrement riche et stimulant. Montréal est une ville qui m'a toujours attiré et je suis heureux d'y avoir vécu deux années complètes, ce qui m'a permis de m'immerger dans sa culture et sa société. J'ai profité de mes moments libres pour commencer l'improvisation théâtrale, m'initier au

Lindy Hop et au Jive ainsi qu'au tennis et j'ai pu profiter des nombreux événements culturels qu'offrent Montréal. Je pense avoir un petit peu exagéré en me rendant trois fois à la cabane à sucre en deux hivers et j'en profite pour remercier le Québec pour son sirop d'érable.

Merci à mes amis de Montréal et de Laval pour tous les moments chaleureux et amicaux que nous avons partagés. Il va sans dire que je remercie énormément ma famille en France et ici pour leur soutien sans failles depuis toujours.

Enfin, j'ai pu réaliser ce projet de mémoire dans des conditions sereines grâce au soutien financier de la Chaire CN en intermodalité (Mai 2013-Août 2014) et je lui en suis très reconnaissant.

Chapitre 1

Introduction

Les réponses apportées aux problèmes ayant trait à l'étude des choix de routes s'appliquent dans la vie des citoyens aussi bien sur le court terme que sur le long terme. Elles offrent la possibilité à des gestionnaires de présenter des arguments scientifiques, avec pour objectif de les aider à prendre des décisions dont l'impact sur la population est potentiellement grand. En effet, des sommes extraordinaires sont engagées dans des défis aussi bien tactiques, comme la régulation d'un trafic urbain, que stratégiques comme la construction d'autoroutes, sans oublier les conséquences au niveau du développement et de la sécurité.

Notre projet étudiera le cas de Contrôle routier Québec (CRQ), qui cherche à optimiser le déploiement de ses ressources de contrôle, sachant que certaines zones du territoire ne sont pas couvertes et qu'il faut renouveler des installations. Les ressources sont composées de sites de contrôles fixes et de patrouilles mobiles. Il y a des contraintes à respecter concernant le budget, le rendement, la couverture du territoire, la sécurité et les lois du travail. La stratégie mise en place vise à augmenter le nombre de contrôles de véhicules en état d'infraction.

Nous proposons comme objectif d'étudier les comportements de ces usagers à travers l'analyse de leurs choix de routes en nous basant sur les modèles de choix discrets. Le cadre théorique présenté dans le chapitre 2, dédié à la revue de littérature est celui des modèles de choix discrets, lesquels nous verrons, ont inspiré des chercheurs dans l'étude de nombreux problèmes dont celui des choix de routes. Ces derniers soulèvent des défis à plusieurs niveaux, que ce soit la difficile collecte d'observations ou la corrélation entre les routes sans oublier les alternatives qui, souvent en nombre élevé et inconnues poussent les analystes à avoir recours à des

stratégies diverses entraînant des lourdeurs et de la difficulté supplémentaire.

Dans le chapitre 3, nous montrons comment nous avons attaqué notre étude. Nous l'avons abordée en commençant par faire une description des données. Cette étape va nous permettre de faire un état des lieux des données et sera l'occasion de rencontrer des personnes qui ont des connaissances précieuses pour nous. Nous faisons une distinction entre les différentes données collectées. Certaines permettront à l'analyste de modéliser le réseau et d'autres, enregistrées lors des contrôles, composeront des observations. La nature exacte de ces dernières sera décrite explicitement. L'étude du comportement des camionneurs peut commencer bien que les observations du terrain soient inexploitablement présentes. Nous avons anticipé ce fait et nous avons eu l'idée d'utiliser des observations simulées qui s'avéreront précieuses lorsque nous aurons besoin de fournir des recommandations concrètes aux personnes qui gèrent les contrôles routiers. Ces recommandations sont des informations qu'il faudra prendre en compte pour des futures collectes de données.

Nous enchaînerons avec le chapitre 4 qui donnera lieu à une présentation des résultats expérimentaux. Pour déterminer le comportement des camionneurs sachant qu'il est modélisé par un modèle de choix discrets, il faut estimer les valeurs des paramètres de la fonction qui donne les probabilités des choix de routes. Alors que dans une situation idéale nous pouvons compter sur des observations complètes en quantité, ici nous devons étudier le comportement des camionneurs avec un ensemble d'observations incomplètes. Les estimations se feront sur 10 échantillons d'observations complètes, simulées grâce à un modèle de choix de routes discrets de référence, puis, nous montrerons comment nous avons réussi à estimer les paramètres avec des échantillons d'observations de plus en plus incomplètes, cela dans une certaine limite, puisque si le manque d'information dépasse un seuil, les estimations n'ont plus de valeur.

Le dernier chapitre prend la forme d'une discussion où on abordera la question de la pertinence de notre approche pour modéliser un comportement et on suggérera des pistes à explorer pour des travaux futurs.

Pour évoquer la contribution de ce mémoire, nous pouvons dire qu'à notre connaissance, alors que le modèle logit récursif a déjà porté ses fruits dans le cas où les observations étaient complètes, c'est la première fois qu'une estimation des paramètres est rendue possible bien que les observations soient partielles. En plus de ce qui a été écrit pour l'utilisation de données synthétiques, nous insistons sur le fait que le re-

cours à ces simulations est une étape incontournable. En effet, connaissant les valeurs des paramètres avant l'estimation, nous mesurons facilement le biais causé par le côté incomplet des observations. Il arrive un moment où celles-ci sont si incomplètes que notre estimation ne sera plus valide. Il faut donc de meilleures observations pour avoir de meilleures inférences. Ceci constitue un argument qui pourra convaincre les gestionnaires de CRQ d'apporter des modifications dans le questionnaire utilisé par les patrouilleurs. En effet, si on rencontre des difficultés dans l'estimation d'un modèle dont on a la maîtrise, alors il est quasi certain que de telles difficultés resurgissent lorsqu'il faudra travailler avec des observations réelles. Nous allons maintenant introduire le fondement théorique des modèles de choix discrets appliqués aux problèmes de choix de routes et procéder à une revue de littérature.

Chapitre 2

Revue de littérature

L'objet de ce chapitre est de passer en revue les diverses contributions scientifiques à l'étude des modèles de choix discrets, de la modélisation à l'estimation dans le contexte particulier des problèmes de choix de routes. L'étude du comportement des usagers sur un réseau présente un intérêt tout aussi bien pour les problèmes liés à la sécurité routière que pour la prédiction des flux de véhicules, sans compter l'aide que cela peut apporter pour la planification des transports, entre autres. Sur un réseau routier, un comportement peut être associé à une façon de prendre des décisions à partir desquelles il sera possible, par exemple, de construire un itinéraire. On peut choisir de circuler dans le réseau en favorisant les chemins qui sont les plus courts; cette stratégie est possible grâce à des applications du web ou du GPS (en anglais pour Global Positioning System), qui est devenu un produit de consommation de masse. La plupart des usagers prennent, par exemple, la décision d'emprunter la route la plus rapide, la moins encombrée ou la plus directe. Certains peuvent préférer rouler en évitant de passer à côté des fumées dégagées par des cheminées d'usines et on peut imaginer d'autres types de comportements encore. En observant des usagers sur un réseau, on peut se demander comment analyser qualitativement les choix de routes.

Une modélisation mathématique est une approche classique pour répondre à ce genre de question. Dans la section 2.1, nous commençons par introduire un cadre théorique, celui des choix discrets, sur lequel repose le modèle de choix de routes que nous proposons. L'idée principale tourne autour de la théorie de l'utilité aléatoire, qui est un concept issu du domaine économique. Cette théorie fait intervenir un décideur considéré rationnel, qui choisit sa route en prenant celle qui a la plus grande

utilité. Il est le seul à savoir la calculer exactement puisqu'il a toutes les informations. Le deuxième intervenant, l'analyste, qui n'a pas tous les éléments pour calculer les utilités, tentera de modéliser le choix de routes du décideur tout en étant contraint d'introduire une variable aléatoire dont il suppose une loi de probabilité. La variable aléatoire matérialise la différence entre ce que sait le décideur et ce qu'observe l'analyste. La nature stochastique qui caractérise le problème des choix de routes du point de vue de l'analyste fait qu'il raisonnera en terme de probabilité de choix de routes, ce que nous expliciterons dans la section 2.1.1.

Après avoir dérivé une formulation générale d'une probabilité de choix de routes pour un modèle logit, nous regarderons une de ses caractéristiques principales dans la section 2.2 et nous décrirons quelques hypothèses pour simplifier la suite de la modélisation dans la section 2.3.

Puis, nous aborderons dans la section 2.4 une méthode d'estimation d'un tel modèle qui donnera l'occasion de résoudre un problème d'optimisation non linéaire et sans contraintes avec une méthode itérative.

Ensuite, nous verrons dans la section 2.5 qu'il existe des difficultés inhérentes aux problèmes de choix de routes, notamment des problèmes combinatoires, mais aussi de corrélation, car non seulement le nombre de routes est difficile, voire impossible à déterminer mais encore, il est quasiment certain que bon nombres d'entre elles partagent des portions communes. Un dernier défi et non des moindres concerne les observations. En effet, avec des observations complètes, l'estimation ne pose pas de problèmes, mais dans les conditions habituelles, les observations sont rares et souvent peu complètes ou bien inexistantes. La section 2.5.4 donnera une piste pour travailler avec des observations incomplètes. Enfin, nous concluons ce chapitre en introduisant dans la section 2.6 le modèle récursif logit de (Fosgerau, Frejinger et Karlström, 2013) qui est le modèle de choix de routes fondamental de cette étude.

2.1 Modèles de choix discrets

Nous commençons par l'introduction d'une classe de modèles mathématiques qui nous permettra d'étudier le comportement des usagers à travers l'analyse de leurs choix de routes. Ce qui suit est fortement inspiré de (Train, 2003) et de (Ben-Akiva et Lerman, 1985) ainsi que de (Frejinger, 2008). L'explication du comportement des usagers sur le réseau peut se faire de multiples manières. Les modèles agrégés,

qui regroupent les preneurs de décision en un seul bloc, et les modèles désagrégés, qui se concentrent sur un preneur de décisions, sont deux façons de procéder. Considérant qu'une décision au niveau collectif est issue de décisions individuelles, et que l'agrégation de données avant estimation d'un modèle peut avoir comme conséquence un manque de précision des estimateurs des moindres carrés ordinaires (Kmenta, 1971), il est raisonnable de privilégier les modèles désagrégés.

L'analyse qualitative des choix peut se faire à l'aide de modèles de choix discrets dans le cadre desquels un preneur de décisions n fait face à un ensemble J_n de choix ou d'alternatives. Cet ensemble non vide doit satisfaire trois critères. D'abord, il doit avoir un nombre fini d'éléments, ensuite, les choix doivent être exhaustifs, et enfin les alternatives sont deux à deux exclusives. Par exemple, un transporteur n de marchandises qui se trouve sur une portion de route décide de prendre au niveau d'une bifurcation une nouvelle portion de route i parmi un ensemble J_n . Il doit faire le choix d'une et une seule portion de route parmi les portions de route en nombre limité, lesquelles lui sont toutes accessibles. Un autre transporteur m peut se trouver à une autre bifurcation et choisir j parmi un ensemble J_m , vraisemblablement différent de J_n . Les ensembles J_n et J_m sont des ensembles de choix discrets, d'où l'appellation des modèles que nous étudions. La nature discrète des alternatives et, par conséquent, l'utilisation des modèles de choix discret est adéquate dans notre analyse de choix de routes. Nous cherchons à déterminer laquelle de ces routes sera choisie par le décideur pour se rendre d'un lieu à un autre. On trouve la première revue de ce problème dans le livre de (Bovy et Stern, 1990) dans lequel il est décrit qu'une route est choisie en fonction de critères associés aux routes (temps de parcours, éclairage, etc...) et en fonction des caractéristiques et des préférences du décideur. Le décideur fait donc un choix et l'analyste va étudier comment il le fait. Dans la prochaine section, nous regarderons plus en détails le processus de choix, en commençant par le côté probabiliste dans une prise de décision et en introduisant la notion centrale de l'utilité. De ce fait, nous allons présenter la théorie de l'utilité aléatoire, qui est une des deux approches pour l'étude de modèles probabilistes.

2.1.1 Théorie de l'utilité aléatoire

Dans (Ben-Akiva et Lerman, 1985), nous voyons que depuis les années soixante, les prises de décision sont expliquées comme étant de nature probabiliste. Dans le

domaine de la psychologie, où ont été menés les premiers travaux sur les théories du choix probabiliste, on s'apercevait que des individus ne prenaient pas systématiquement la même décision alors qu'on leur présentait constamment les mêmes choix possibles. Dans (Luce et Suppes, 1965), les auteurs passent en revue les développements sur ce qu'on qualifie de théorie du choix probabiliste. Dans la théorie de l'utilité aléatoire, idée d'abord apparue dans (Thurston, 1927) puis formellement décrite dans (Manski, 1977), nous avons deux intervenants, le preneur de décisions et l'analyste.

Le choix est fait par le preneur de décisions, celui-ci peut être une personne, un groupe de personnes ou une entité (un transporteur routier de marchandises, un comité d'administration, un ministère, etc...). Parfois, on doit faire abstraction du groupe et l'assimiler à un décideur unique parce qu'on ne connaît pas les interactions entre les membres de l'unité de décisions. Dans ce qui suit, le preneur de décisions sera une seule personne, appelée le décideur. Il sait calculer l'utilité de chaque alternative et il choisit celle qui a la plus grande valeur, ce qui fait que son choix est déterministe.

L'analyste, chargé de la modélisation, peut à partir de données (lorsqu'il y en a) sur le preneur de décisions n et sur les alternatives, prédire le comportement du décideur n en calculant la probabilité $\Pr_n(i)$ qu'il choisisse une alternative i dans J_n . Cet aspect stochastique est dû au fait que l'analyste est obligé de considérer une variable aléatoire, reflétant son manque d'information, mais il peut émettre des hypothèses sur la loi de probabilité de cette variable aléatoire et nous verrons que la nature du modèle en dépendra. Les auteurs de (Burrell, 1968) et (Dial, 1971) sont parmi les premiers à traiter des problèmes de choix de routes comme des problèmes de nature stochastique en utilisant la théorie de l'utilité aléatoire.

Nous allons aborder la notion d'utilité et dériver une expression de la probabilité de choisir une alternative.

2.1.2 Dérivation de la probabilité de choix d'une alternative par un décideur

Notons que le décideur fait ses choix en se basant sur les attributs des alternatives s'offrant à lui et en fonction de critères socio-économiques qui lui sont propres. Les attributs des alternatives sont en général perçus différemment d'un décideur à l'autre.

En pratique, l'analyste dispose de données loin d'être complètes et totalement fiables. Lorsqu'un décideur choisit une portion de route, l'analyste peut observer des attributs comme la longueur de cette portion et la vitesse maximale y étant autorisée. Parfois, il peut avoir à sa disposition l'âge du conducteur, mais d'autres attributs tels que la proximité sur la portion de route à choisir d'une usine polluante, le revenu du conducteur ou l'heure d'un contrôle routier peuvent lui échapper, alors que ce sont potentiellement des données pertinentes pour expliquer un comportement. En notant x_{ni} et s_n les attributs observables par l'analyste, d'une alternative i et du décideur n respectivement, la probabilité que n choisisse i dans J_n s'écrit:

$$\Pr_n(i) = g(s_n, x_{ni} \forall i \in J_n, \beta) \quad (2.1)$$

Le calcul des probabilités va dépendre de la façon dont l'analyste va spécifier son modèle, à savoir comment il va définir la fonction g . Le vecteur de paramètres β est à déterminer et est un vecteur colonne de scalaires de taille $Q \geq 1$, lorsqu'il y a Q attributs observables. Chaque attribut observable se voit attribuer un poids, qui est un scalaire du vecteur β . Pour le moment, nous considérons que β ne dépend ni de n , ni des alternatives. Déterminer β revient à déterminer le modèle expliquant les choix des décideurs.

Nous avons supposé que le décideur est capable de calculer une utilité pour chaque alternative avant de faire son choix. L'utilité est une des quatre catégories de règles de prise de décision, les autres étant la dominance, la satisfaction et la lexicographie selon (Ben-Akiva et Lerman, 1985). Elle stipule, qu'à partir d'un vecteur donné d'attributs, on peut calculer un scalaire, qui représente l'attractivité de l'alternative associée à ces attributs. Les problèmes que nous étudierons seront ceux où il faudra maximiser la fonction d'utilité. Le comportement rationnel du décideur, outre le fait qu'il cherche à maximiser l'utilité qui découle de son choix, est caractérisé par le fait que le calcul de l'utilité repose sur des critères objectifs. L'alternative $i \in J_n$ procure une utilité U_{ni} pour le décideur n , sachant qu'il ne prend en compte que les attributs z_{ni} de i qui lui sont pertinents. De même, on notera r_n les attributs pertinents du décideur n pris en compte par lui-même. Cette utilité s'écrit:

$$U_{ni} = U(z_{ni}, r_n), \forall i \in J_n \quad (2.2)$$

U étant une fonction à valeurs réelles.

Le décideur choisit l'alternative $i \in J_n$ si et seulement si:

$$U_{ni} > U_{nj}, \forall j \neq i$$

ce qui équivaut à , n choisit $i \in J_n$ si et seulement si:

$$U(z_{ni}, r_n) > U(z_{nj}, r_n), \forall j \neq i.$$

Ici, le décideur fait son choix de manière déterministe, puisque si i a la plus grande utilité, sa probabilité d'être choisie est 1. L'aspect stochastique vient du fait que l'analyste ne connaît ni les attributs pertinents, ni la fonction d'utilité U . Sa connaissance de la fonction d'utilité se limite à la représentation qu'il s'en fait à l'aide des attributs observés, à savoir, x_{ni} et s_n , qui seront les variables d'une fonction V_{ni} . Ce différentiel de connaissances entre le décideur et l'analyste constitue une part importante dans ce que (Manski, 1977) nomme la théorie de l'utilité aléatoire. Dans cette théorie, l'utilité $U_{ni}(z_{ni}, r_n)$ se décomposera en deux parties, $V_{ni}(x_{ni}, s_n, \beta)$, la partie déterministe de l'utilité, calculable lorsque le vecteur de coefficient β est déterminé, et de ϵ_{ni} , la partie non observable de l'utilité. Cette dernière, simplement définie comme la différence entre l'utilité procurée au décideur et l'utilité calculée par l'analyste, sera traitée comme une variable aléatoire.

$$U_{ni} = V_{ni}(x_{ni}, s_n, \beta) + \epsilon_{ni}, \forall i \tag{2.3}$$

La densité de distribution est $f(\epsilon_{nj})$ et la fonction de répartition est $F(\epsilon_{nj})$. Le choix d'une densité va influencer sur la formulation du modèle de choix de route. En effet, si nous supposons que les erreurs suivent une loi normale jointe, alors notre modèle suivra une formulation probit. Les termes d'erreur sont définis par une matrice de variance-covariance, ce qui suggère que la corrélation entre deux termes d'erreur peut être non nulle et que leurs variances ne sont pas nécessairement les mêmes. Le modèle probit est néanmoins difficile à utiliser, car la densité des termes d'erreur s'exprime à l'aide d'une intégrale. C'est le prix à payer si l'analyste veut par exemple étudier un modèle dans lequel les termes d'erreur peuvent être dépendants les uns des autres et distribués de manière non identique, pour gagner en réalisme. Dans (Ben-Akiva et Lerman, 1985), il est précisé que le modèle probit multinomial, bien qu'introduit il y a plusieurs décennies par (Thurston, 1927), a montré ses premiers

résultats avec les travaux de (Dutt, 1976), (Hausman et Wise, 1978), (Daganzo, Bouthelie et Sheffi, 1977) puis (Albright, Manski et Lerman, 1977) qui ont résolu des problèmes computationnels. Ensuite, ce modèle n'est que très marginalement utilisé et il existe très peu de travaux concernant les problèmes de demande de transport l'utilisant.

Une autre façon de modéliser un choix de routes est le modèle logit dans lequel nous pouvons écrire de manière plus simple $f(\epsilon_{nj})$, la densité de la loi de probabilité et $F(\epsilon_{nj})$, la fonction de répartition avec chaque ϵ_{nj} qui suit une loi d'extremum de type I, indépendamment et identiquement distribuée (iid) (aussi appelée loi de Gumbel). Contrairement à ce qui a été vu pour le modèle probit, cette fois, les termes d'erreur ont des variances égales et les covariances entre deux termes d'erreur sont nulles.

$$f(\epsilon_{nj}) = \frac{1}{\mu} e^{-\frac{\epsilon_{nj}-\lambda}{\mu}} e^{-e^{-\frac{\epsilon_{nj}-\lambda}{\mu}}} \quad (2.4)$$

$$F(\epsilon_{nj}) = e^{-e^{-\frac{\epsilon_{nj}-\lambda}{\mu}}} \quad (2.5)$$

Le paramètre de localisation λ est à valeurs réelles. Le facteur d'échelle μ permet de rendre compte de la variance des erreurs sur la valeur des paramètres du modèle à estimer. Dans le cas standard, $\lambda = 0$ et $\mu = 1$, la variance de la distribution du terme d'erreur est $\frac{\pi^2}{6}$ et son espérance $\gamma \sim 0,577$, la constante d'Euler. A première vue, l'hypothèse d'indépendance des erreurs est restrictive. Elle implique que l'erreur pour une alternative ne renseigne en rien sur l'erreur pour une autre alternative. Dans le cas où l'analyste représente l'utilité très justement, le terme d'erreur a les propriétés d'un bruit blanc. Le modèle logit est alors approprié. En effet, utiliser la loi de Gumbel pour les erreurs ϵ_{ni} est pratiquement équivalent à dire que les erreurs sont indépendantes entre-elles et de loi normale identique $N(0, \mu^2)$. Dans (Ben-Akiva et Lerman, 1985), la loi de Gumbel peut être vue comme une approximation de la loi normale et a l'avantage de mener à des calculs plus simples. On peut calculer

la probabilité que le preneur de décisions n choisisse l'alternative i qui est selon (McFadden, 1974):

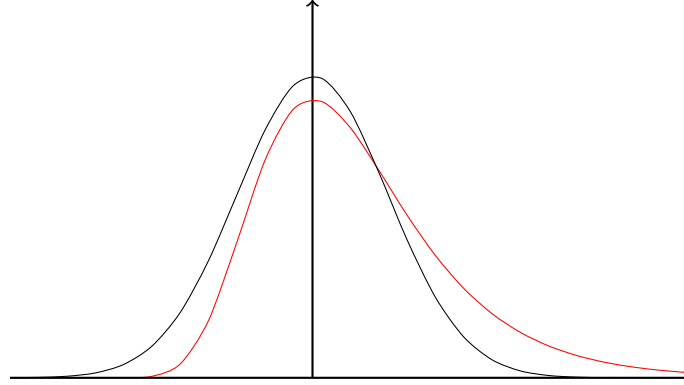
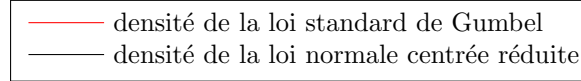


Figure 2.1 – La loi de Gumbel comme approximation de la loi normale centrée réduite

$$\Pr_n(i) = \Pr(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj}, \forall j \neq i) \quad (2.6)$$

$$= \Pr\left(\frac{V_{ni} + \epsilon_{ni}}{\mu} > \frac{V_{nj} + \epsilon_{nj}}{\mu}, \forall j \neq i\right), \mu > 0, \quad (2.7)$$

$$= \Pr\left(\frac{\epsilon_{nj}}{\mu} < \frac{\epsilon_{ni} + V_{ni} - V_{nj}}{\mu}, \forall j \neq i\right), \mu > 0, \quad (2.8)$$

$$= \Pr(\check{\epsilon}_{nj} < \check{\epsilon}_{ni} + \frac{V_{ni} - V_{nj}}{\mu}, \forall j \neq i), \mu > 0 \quad (2.9)$$

en posant $\check{\epsilon}_{nj} = \frac{\epsilon_{nj}}{\mu}$, $\forall j \in J_n$ et $\mu > 0$.

La démonstration reprend celle de (Train, 2003). La fonction de répartition de chaque variable aléatoire $\check{\epsilon}_{nj}$ exprimée en $\check{\epsilon}_{ni} + \frac{V_{ni} - V_{nj}}{\mu}$ est, en prenant $\lambda = 0$ pour simplifier d'après 2.5:

$$\exp(-\exp(-(\check{\epsilon}_{ni} + \frac{V_{ni} - V_{nj}}{\mu})))$$

Avec l'hypothèse que les termes d'erreur ϵ sont indépendants, la fonction de répartition pour tout $j \neq i$ est le produit des fonctions de répartition individuelles:

$$\Pr_n(i)|_{\check{\epsilon}_{ni}} = \prod_{j \neq i} e^{-e^{-(\check{\epsilon}_{ni} + \frac{V_{ni} - V_{nj}}{\mu})}} \quad (2.10)$$

Ceci a été calculé en supposant $\check{\epsilon}_{ni}$ connus, ce qui n'est pas le cas. Il faut alors intégrer $\Pr_n(i)|_{\check{\epsilon}_{ni}}$ par rapport à $\check{\epsilon}_{ni}$ sur toutes ses valeurs en pondérant par sa densité de distribution pour avoir une expression de $\Pr_n(i)$, ce qui donne:

$$\Pr_n(i) = \int_{\check{\epsilon}_{ni}=-\infty}^{\infty} \left(\prod_{j \neq i} e^{-e^{-\check{\epsilon}_{ni} + \frac{V_{ni} - V_{nj}}{\mu}}} \right) e^{-\check{\epsilon}_{ni}} e^{-e^{-\check{\epsilon}_{ni}}} d\check{\epsilon}_{ni}$$

En remarquant que $V_{ni} - V_{ni} = 0$, on peut regrouper les termes en $e^{-e^{-\check{\epsilon}}}$ sous la même parenthèse et on n'a plus la restriction $j \neq i$, ce qui donne:

$$\begin{aligned} \Pr_n(i) &= \int_{\check{\epsilon}_{ni}=-\infty}^{\infty} \left(\prod_{j \in J_n} e^{-e^{-\left(\check{\epsilon}_{ni} + \frac{V_{ni} - V_{nj}}{\mu}\right)}} \right) e^{-\check{\epsilon}_{ni}} d\check{\epsilon}_{ni} \\ &= \int_{\check{\epsilon}_{ni}=-\infty}^{\infty} \exp\left(-\sum_{j \in J_n} e^{-\left(\check{\epsilon}_{ni} + \frac{V_{ni} - V_{nj}}{\mu}\right)}\right) e^{-\check{\epsilon}_{ni}} d\check{\epsilon}_{ni} \\ &= \int_{\check{\epsilon}_{ni}=-\infty}^{\infty} \exp\left(-e^{-\check{\epsilon}_{ni}} \sum_{j \in J_n} e^{-\left(\frac{V_{ni} - V_{nj}}{\mu}\right)}\right) e^{-\check{\epsilon}_{ni}} d\check{\epsilon}_{ni} \end{aligned}$$

Avec le changement de variable $t = e^{-\check{\epsilon}_{ni}}$, nous avons $dt = -e^{-\check{\epsilon}_{ni}} d\check{\epsilon}_{ni}$. Nous avons donc:

$$\begin{aligned} \Pr_n(i) &= \int_{\infty}^0 \exp\left(-t \sum_{j \in J_n} e^{-\frac{1}{\mu}(V_{ni} - V_{nj})}\right) (-dt) \\ &= \int_0^{\infty} \exp\left(-t \sum_{j \in J_n} e^{-\frac{1}{\mu}(V_{ni} - V_{nj})}\right) dt \\ &= \left[\frac{\exp\left(-t \sum_{j \in J_n} e^{-\frac{1}{\mu}(V_{ni} - V_{nj})}\right)}{-\sum_{j \in J_n} e^{-\frac{1}{\mu}(V_{ni} - V_{nj})}} \right]_0^{\infty} \\ &= \frac{1}{\sum_{j \in J_n} e^{-\frac{1}{\mu}(V_{ni} - V_{nj})}} \end{aligned}$$

Finalement, la probabilité dans le modèle de choix logit s'écrit:

$$\Pr_n(i) = \frac{e^{\frac{1}{\mu}V_{ni}}}{\sum_{j \in J_n} e^{\frac{1}{\mu}V_{nj}}} > 0 \quad (2.11)$$

$$\sum_{i \in J_n} \Pr_n(i) = \frac{\sum_{i \in J_n} e^{\frac{1}{\mu}V_{ni}}}{\sum_{j \in J_n} e^{\frac{1}{\mu}V_{nj}}} = 1$$

La somme des probabilités vaut 1 comme attendu et $\Pr_n(i)$ est strictement positif. Plus l'utilité déterministe de i tend vers $-\infty$, plus $\Pr_n(i)$ tend vers 0. Plus utilité de i

croit, plus $\text{Pr}_n(i)$ tend vers 1 avec les V_{nj} maintenus constants. Maintenant que nous avons une expression du choix de probabilité pour un modèle logit, regardons une de ses propriétés caractéristiques.

2.2 La propriété IIA du logit

Une propriété remarquable du modèle logit est l'indépendance des alternatives non pertinentes (Independence of Irrelevant Alternatives, IIA en anglais). On remarque que le ratio $\frac{\text{Pr}_n(i)}{\text{Pr}_n(j)} = e^{\frac{1}{\mu}(V_{ni}-V_{nj})}$ ne dépend que de $\frac{1}{\mu}(V_{ni} - V_{nj})$, c'est à dire qu'une alternative k , $k \neq i, j$, n'est pas pertinente dans le choix entre les alternatives i ou j .

Cette propriété permet au modèle d'être estimé en utilisant seulement un échantillon de l'ensemble des alternatives J_n , ce qui a pour conséquence des calculs moins coûteux (McFadden, 1978). On peut s'inspirer de (Train, 2003), si on considère que le décideur n a le choix entre 100 routes toutes connues, on peut faire l'estimation du modèle à un moindre coût sur 10 routes, dont une est choisie par le décideur et les 9 autres sont tirées aléatoirement parmi les 99 routes restantes. Dans ce cas, l'IIA assure que les ratio des probabilités sont constants et ne dépendent pas des attributs, ni de l'existence de routes non présentes dans l'échantillon et en fin de compte l'estimateur converge vers la même valeur quelque soit l'échantillon. Il s'agit toutefois d'un exemple assez éloigné de la pratique pour les problèmes de choix de routes puisque l'analyste ne connaît pas les 100 routes, aussi il n'est pas garanti d'avoir la route observée dans son échantillon et par conséquent, les estimateurs ne convergent pas vers la même valeur suivant l'échantillon. Citons (Train, 2003): "Dans de nombreux cas, les modèles tels que le logit qui présentent la propriété de l'IIA offrent une modélisation fine de la réalité. Dans (Luce, 1959), l'auteur voyait dans l'IIA une propriété caractéristique des modèles bien spécifiés".

Cependant, la propriété de l'IIA peut mener à des situations peu réalistes. Lorsqu'une nouvelle alternative similaire à une autre déjà dans J_n est ajoutée, dans certains cas les calculs donnent des résultats paradoxaux. Un exemple inspiré du célèbre cas autobus bleu et autobus rouge pour illustrer ceci est le suivant: imaginons un usager faisant face à un choix entre deux routes pour aller travailler. Il a le choix entre prendre une première route directe goudronnée et sans défauts, et une autre composée de deux morceaux dont la qualité de revêtement est très inégale. Le premier morceau est goudronné et sans défauts alors que le deuxième morceau est de très

mauvaise qualité avec un nombre $nb1$ de nids de poules. Supposons que la probabilité qu'il décide de prendre la première route ($route_1$) est deux fois plus élevée que celle de prendre la deuxième ($route_2$), c'est-à-dire $\frac{\Pr(route_1)}{\Pr(route_2)} = 2$ et selon la propriété de l'IIA, ce ratio reste constant. En respectant les propriétés des calculs de probabilité, ces deux probabilités sont $\Pr(route_1) = \frac{2}{3}$ et $\Pr(route_2) = \frac{1}{3}$. Maintenant, ajoutons un troisième choix, quasiment identique au deuxième, qui est une route ($route_3$) composée de deux morceaux dont le premier est le même que le premier morceau de la $route_2$ et le deuxième morceau est similaire à celui de la $route_2$ mais avec un nombre $nb2$ de nids de poule. Autrement dit, la $route_3$ ainsi ajoutée à l'ensemble des alternatives du décideur est quasiment identique à une alternative présente auparavant. Le décideur est supposé choisir en fonction de la qualité du revêtement sans tenir compte du nombre de nids de poule sur la route.

La propriété de l'IIA nous dit que la probabilité de choisir la $route_1$ reste deux fois plus élevée par rapport à celle de choisir la $route_2$ malgré l'introduction de la $route_3$ dans l'ensemble des alternatives. Dans cet exemple on aura alors $\Pr(route_1) = \frac{1}{2}$ et $\Pr(route_2) = \Pr(route_3) = \frac{1}{4}$.

Or, le décideur perçoit les $route_2$ et $route_3$ de la même façon, donc $\Pr(route_1) = \frac{2}{3}$ et $\Pr(route_2 \cup route_3) = \frac{1}{3}$ et puisque celles-ci ont la même probabilité d'être choisies et ne peuvent pas être choisies en même temps, alors $\Pr(route_2) = \Pr(route_3) = \frac{1}{6}$.

Si la propriété de l'IIA se maintient, la probabilité de prendre la $route_1$, de bonne qualité partout sera sous-estimée alors que les probabilités de prendre les routes 2 et 3 partiellement dégradées seront surestimées. Nous savons que la propriété de l'IIA est caractéristique des modèle de type logit, car elle est une conséquence de l'hypothèse selon laquelle les erreurs sont iid. Un moyen de passer outre les paradoxes induits par l'IIA est, par exemple, d'utiliser un modèle de type probit, mais nous avons vu que le modèle probit est compliqué à manier et coûteux à cause de sa formulation par intégrale. D'autres observations peuvent être faites sur la formule du logit. Nous avons vu par exemple que le facteur d'échelle μ est arbitraire (mais strictement positif) dans la dérivation de $\Pr_n(i)$. Nous montrons dans la prochaine partie quelques approximations qui vont simplifier notre modélisation.

2.3 Approximations

2.3.1 Homoscédasticité

Le facteur d'échelle est présent dans tous les arguments des termes en exponentielle et on supposera que c'est une constante. Seul le rapport entre l'utilité déterministe et le facteur d'échelle sera estimable. De plus, cela sera impossible d'identifier le dernier seul lors d'une estimation. Pour le reste de l'étude, nous fixons sa valeur à 1 pour simplifier. Jusqu'à présent, nous avons défini un modèle où le vecteur β est le même pour toute la population, ce qui implique que tous les individus ont le même comportement. On peut s'attendre à ce que la population soit composée de groupes distincts g socio-économiques homogènes caractérisés par leur propres vecteurs β_g . C'est le principe de la segmentation de marché. Dans (Ben-Akiva et Lerman, 1985), il est dit qu'une segmentation de marché peut résulter dans la détection d'hétéroscédasticité malgré des valeurs de paramètres égales. Le facteur d'échelle rend compte de la variance des erreurs des utilités; si, dans la population, cette hypothèse d'homoscédasticité ne tient plus, nous devons affiner notre modèle afin qu'il soit plus réaliste.

2.3.2 Cas limites

Rappelons que, pour une loi de Gumbel de paramètres $\lambda = 0$ et $\mu > 0$, sa variance est de $\frac{\mu^2 \pi^2}{6}$. L'analyste doit faire attention quand il fixe une valeur arbitraire de μ , car le ratio entre la variance et les mesures des utilités déterministes a une influence sur la nature du comportement. Si on considère l'expression de la loi logit 2.11, nous avons deux cas limites évoqués dans (Ben-Akiva et Lerman, 1985):

1. Cas 1:

$$\lim_{\mu \rightarrow \infty} \Pr_n(i) = \frac{1}{|J_n|}, \forall i \in J_n$$

La variance tendant vers l'infini, le modèle est défini tel que les erreurs l'emportent sur l'utilité observée et n'apporte donc aucune information. Les alternatives sont choisies de manière équiprobable.

2. Cas 2:

$$\lim_{\mu \rightarrow 0} \Pr_n(i) = \lim_{\mu \rightarrow 0} \frac{e^{\frac{1}{\mu} V_{ni}}}{e^{\frac{1}{\mu} V_{ni}} + \sum_{\substack{j \in J_n \\ j \neq i}} e^{\frac{1}{\mu} V_{nj}}}, \forall i \in J_n$$

en divisant le numérateur et le dénominateur par $e^{\frac{1}{\mu}V_{ni}}$, nous avons:

$$\lim_{\mu \rightarrow 0} \Pr_n(i) = \lim_{\mu \rightarrow 0} \frac{1}{1 + \sum_{\substack{j \in J_n \\ j \neq i}} e^{\frac{V_{nj} - V_{ni}}{\mu}}}, \forall i \in J_n$$

$$= \begin{cases} 1 & \text{si } V_{ni} > \max_{\substack{j \in J_n \\ j \neq i}} V_{nj}, \\ 0 & \text{si } V_{ni} < \max_{\substack{j \in J_n \\ j \neq i}} V_{nj} \end{cases}$$

Il se peut que l'on ait $V_{ni} = \max_{j \in J_n} V_{nj}$ pour un ensemble de valeurs $j \in J_n^* \subset J_n$. Pour cet ensemble J_n^* , la limite vaut $\frac{1}{|J_n^*|}$ et vaut 0 pour l'ensemble des alternatives dans $J_n - J_n^*$. Lorsque le facteur d'échelle tend vers 0, la variance des erreurs tend vers 0. Dans ce cas, tout se passe comme si la fonction d'utilité n'a plus de variable aléatoire et le modèle logit tend vers un modèle de choix déterministe.

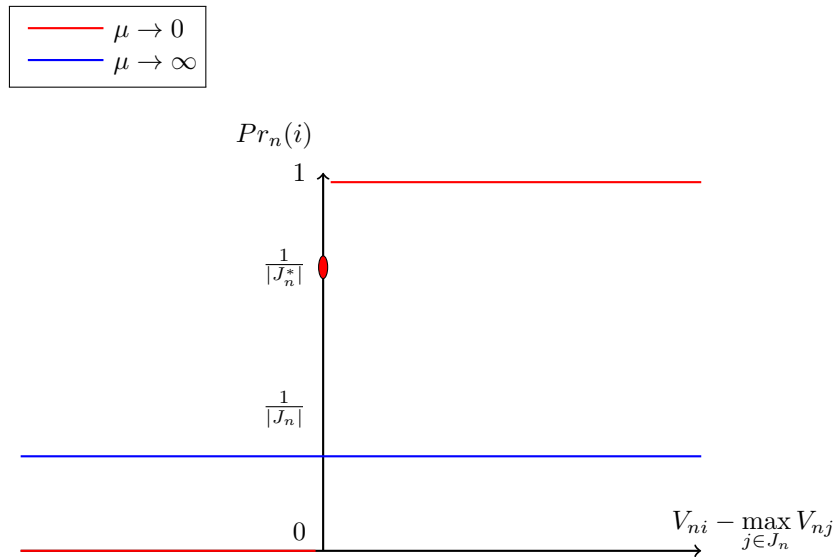


Figure 2.2 – Cas limites

En pratique, nous arrivons bien à mettre en évidence ces deux comportements et dans la suite du projet nous veillerons à ce que les valeurs numériques des termes d'erreurs ne soient ni nulles ni prépondérantes par rapport aux

valeurs des termes d'utilité déterministes, afin d'éviter de se retrouver avec des processus de prise de décision totalement déterministes ou totalement aléatoires.

2.3.3 Une fonction d'utilité linéaire par rapport aux paramètres

L'analyste fait l'hypothèse qu'il peut représenter l'utilité grâce à une fonction d'utilité déterministe dont il connaît la forme. Supposons que la fonction V_{ni} soit linéaire pour le vecteur β , alors (2.11) s'écrit:

$$\Pr_n(i) = \frac{e^{\frac{1}{\mu}\beta^T(x_{ni}^T, s_n^T)^T}}{\sum_{j \in J_n} e^{\frac{1}{\mu}\beta^T(x_{nj}^T, s_n^T)^T}} \quad (2.12)$$

et pour alléger l'écriture, nous posons $x_{ni} = (x_{ni}^T, s_n^T)^T$ où y^T est la transposée d'un vecteur y^T quelconque. Ce qui donne:

$$\Pr_n(i) = \frac{e^{\frac{1}{\mu}\beta^T x_{ni}}}{\sum_{j \in J_n} e^{\frac{1}{\mu}\beta^T x_{nj}}} \quad (2.13)$$

Cette hypothèse simplificatrice peut se justifier par le fait que nous pouvons approximer une fonction par son développement de Taylor au premier ordre, pourvu qu'elle soit différentiable. Une fois un modèle proposé, nous pouvons passer à la prochaine étape qui consiste à estimer les paramètres β .

2.4 Estimation d'un modèle logit

2.4.1 Le maximum de vraisemblance

Dans le cas d'un modèle logit, dont la formulation est relativement simple, la méthode de l'estimation du maximum de vraisemblance est la plus répandue car l'estimateur du maximum de vraisemblance, qui est la variable aléatoire $\widehat{\beta}_N$, présente des caractéristiques intéressantes asymptotiquement (N étant le nombre d'observations dans l'échantillon).

Dans (Ben-Akiva et Lerman, 1985), les auteurs mentionnent les caractéristiques asymptotiques de cet estimateur:

1. Pas de biais: $\lim_{N \rightarrow \infty} E[\widehat{\beta}_N] = \beta$, β étant la valeur déterministe des paramètres pour la population.
2. Convergence en probabilité: $\lim_{N \rightarrow \infty} \Pr(\beta - q \leq \widehat{\beta}_N \leq \beta + q) = 1$ pour tout q positif et arbitrairement petit.
3. Normalité: la loi de $\widehat{\beta}_N$ tend vers une loi normale.
4. Efficacité: si $\widehat{\beta}_N$ converge en probabilité, alors la matrice Σ_{cov} est définie positive: $\Sigma_{cov} = \Sigma_{\widehat{\theta}_N} - \Sigma_{\widehat{\beta}_N}$ où $\Sigma_{\widehat{\theta}_N}$ et $\Sigma_{\widehat{\beta}_N}$ sont les matrices de covariance de $\widehat{\beta}_N$ et de tout autre estimateur convergent en probabilité vers β respectivement.

Supposons que nous disposions d'un échantillon de N preneurs de décisions ayant chacun choisi une alternative. La probabilité qu'un preneur de décisions n choisisse l'alternative que l'analyste a observée s'écrit:

$$\prod_{i \in J} (\Pr_n(i, \beta))^{y_{ni}} \quad (2.14)$$

avec $y_{ni} = 1$ si n choisit i et 0 sinon. J est l'ensemble des alternatives. En supposant que les preneurs de décisions font leur choix indépendamment les uns des autres, alors la probabilité que tous les preneurs de décisions choisissent l'alternative que l'analyste a en effet observée pour chacun d'eux est:

$$L(\beta) = \prod_{n=1}^N \prod_{i \in J} (\Pr_n(i, \beta))^{y_{ni}} \quad (2.15)$$

On dit que $L(\beta)$ est la fonction de vraisemblance (likelihood en anglais). L'estimation se fera plutôt avec le logarithme de la vraisemblance (ou log-vraisemblance) $LL(\beta)$ pour simplifier les calculs et cela ne changera pas la valeur des paramètres estimés étant donné que la fonction logarithme est strictement croissante sur son domaine de définition. Faire l'usage du logarithme de la vraisemblance se révélera être une solution payante puisque dans un modèle logit, nous dérivons une expression simple qui fait intervenir linéairement les valeurs des attributs. En effet, d'après les expressions 2.13 et 2.15:

$$LL(\beta) = \sum_{n=1}^N \sum_{i \in J} y_{ni} \left(\beta^T x_{ni} - \ln \sum_{j \in J_n} e^{\beta^T x_{nj}} \right) \quad (2.16)$$

Non seulement le passage par le logarithme peut simplifier les fonctions en jeu, mais aussi, il permet d'éviter des problèmes numériques liés à la capacité des ma-

chines à représenter des nombres trop grands ou trop petits. Ce problème de virgule flottante peut arriver quand la taille de l'échantillon augmente.

Dans l'étude (McFadden, 1974), l'auteur démontre que $LL(\beta)$ pour une fonction logit multinomiale est globalement concave lorsque la fonction utilité est linéaire en β , en rappelant que $\beta = [\beta_q]$, un vecteur colonne de dimension $Q \geq 1$.

L'estimateur du maximum de vraisemblance (EMV), qui permet d'obtenir le maximum global de la fonction de vraisemblance, elle-même continue et différentiable par rapport à β_q , est le vecteur $\hat{\beta} = \{\hat{\beta}_q\}$, $\forall q \in [1, Q]$ tel que:

$$\frac{\partial LL(\beta)}{\partial \beta_q}(\hat{\beta}_q) = 0, \forall q \in [1, Q] \quad (2.17)$$

D'après (2.16), cette équation devient:

$$\frac{\partial LL(\beta)}{\partial \beta_q}(\hat{\beta}_q) = \sum_{n=1}^N \sum_{i \in J} y_{in}(x_{inq} - \frac{\sum_{j \in J_n} x_{j n q} e^{\hat{\beta}^T x_{nj}}}{\sum_{j \in J_n} e^{\hat{\beta}^T x_{nj}}}) = 0, \forall q \in [1, Q] \quad (2.18)$$

x_{inq} est la valeur de l'attribut q en considérant l'alternative i choisie par n , et finalement:

$$\frac{\partial LL(\beta)}{\partial \beta_q}(\hat{\beta}_q) = \sum_{n=1}^N \sum_{i \in J} [y_{in} - \text{Pr}_n(i)] x_{inq} = 0, \forall q \in [1, Q] \quad (2.19)$$

constitue la condition au premier ordre, qui est nécessaire et suffisante.

Pour estimer le maximum de la fonction de vraisemblance, il existe des méthodes bien connues telles que la méthode du gradient et la méthode de Newton-Raphson, surtout utilisées en théorie, car elles demandent des calculs exacts de gradients et de hessiens, respectivement. Nous choisirons d'utiliser la méthode de Broyden-Fletcher-Goldfarb-Shanno (BFGS) vue dans (Train, 2003), combinée à une recherche de direction de descente, permettant de résoudre des problèmes d'optimisation non linéaires sans contraintes, qui présente des avantages par rapport aux deux premières méthodes. Le travail fait dans (Mai-Anh, 2012) a été d'une grande aide pour cette partie.

2.4.2 Résolution d'un problème d'optimisation non linéaire sans contraintes

Le point fort de la méthode de Newton-Raphson est qu'elle permet de trouver une direction d'amélioration à chaque itération lorsque la fonction à maximiser est globalement concave. En contrepartie, elle fait intervenir un calcul coûteux du hessien à chaque itération et est donc peu utilisée en pratique. Ce qui intéresse l'analyste, c'est une méthode qui tient compte des particularités de son problème d'optimisation, qui n'est pas trop lourde à implémenter sans oublier une convergence la plus rapide possible en partant d'un point initial arbitraire ¹.

La méthode BFGS est une méthode de type quasi-Newton où, à chaque itération, une approximation du hessien est calculée, appelée arc hessien. Il s'obtient grâce à la valeur du gradient en deux points qui ne sont pas dans un voisinage l'un de l'autre puisque l'on cherche à déplacer β_t à un niveau non plus infinitésimal mais macroscopique. Le gradient peut être calculé analytiquement, ce qui rendra plus lourde l'implémentation, ou encore par la méthode des différences finies et on obtient alors une approximation. L'arc hessien représente la variation de gradient lors d'un déplacement sur le graphe de la fonction de log-vraisemblance, de $LL(\beta_t)$ vers $LL(\beta_{t+1})$.

Avant d'utiliser une méthode de descente, on peut dans un premier temps formuler notre problème d'optimisation en un problème de minimisation, puisque rechercher un maximum d'une fonction revient à chercher le minimum pour la fonction de signe opposé. Notre problème d'optimisation s'écrit:

$$\min_{\beta \in \mathbb{R}^Q} (-LL(\beta)) \tag{2.20}$$

1. voir discussion sur la résolution de l'équation de Bellman dans la section 2.6

Un algorithme de recherche de l'(EMV) peut s'écrire:

Données : $\beta_0 = \beta_{ini}$, $B_0 = I$ et $t = 0$;

Résultat : une valeur de l'EMV $\hat{\beta}$

initialisation

tant que (*condition(s) d'arrêt non atteinte(s)*) **faire**

calculer la direction de descente d_t telle que :

$$B_t d_t = -\nabla LL(\beta_t);$$

calculer la longueur optimale du pas α_t par recherche linéaire dans la direction d_t puis effectuer

$$\beta_{t+1} = \beta_t + \alpha_t d_t;$$

et

$$t \leftarrow t + 1;$$

mettre à jour B_t (voir 2.23) ;

fin

Algorithme 1 : Méthode de descente combinée au BFGS

où B_t est l'arc hessien, d_t la direction de descente, $\nabla LL(\beta_t)$ le gradient de la log-vraisemblance à l'itération t sachant que la mise à jour de B_t satisfait l'approximation de la sécante:

$$B_{t+1}(\beta_{t+1} - \beta_t) = \nabla LL(\beta_{t+1}) - \nabla LL(\beta_t) \quad (2.21)$$

En posant $s_t = (\beta_{t+1} - \beta_t)$ et $y_t = \nabla LL(\beta_{t+1}) - \nabla LL(\beta_t)$, nous avons:

$$B_{t+1} s_t = y_t \quad (2.22)$$

L'approximation du hessien dans la méthode BFGS est telle que:

$$B_{t+1} = B_t - \frac{B_t s_t s_t^T B_t}{s_t^T B_t s_t} + \frac{y_t y_t^T}{y_t^T s_t} \quad (2.23)$$

La méthode pour trouver la valeur optimale $\hat{\beta}$ fait appel à un algorithme qui, à partir d'un vecteur initial, génère une suite convergente de vecteurs. L'algorithme que nous avons présenté fait appel à une méthode qui recherche une direction de descente et un pas tout en respectant les critères de Wolfe. Le but n'est pas la recherche d'une solution exacte à chaque itération, ce qui pourrait prendre trop de temps. Il s'agit de trouver une approximation de la valeur du pas pour la direction de descente courante qui fasse suffisamment descendre la valeur de la fonction objectif. À chaque itération, on cherche donc α tel que:

$$\min_{\alpha > 0} h(\beta_t + \alpha d_t)$$

avec $h(\cdot) = -LL(\cdot)$.

Une fois α trouvé, une nouvelle direction de descente est calculée et on met à jour l'arc hessien .

On refait des itérations ainsi de suite jusqu'à ce qu'au moins un critère d'arrêt soit satisfait. La recherche de l'optimum s'arrête quand la norme du gradient de $LL(\beta_t)$ est inférieure à une valeur seuil; dans ce cas, il y a convergence. Une autre raison d'arrêter de chercher l'optimum est lorsque la valeur de $LL(\beta_t)$ reste sensiblement la même en étant supérieure en valeur absolue à la valeur seuil alors que le nombre d'itérations ne cesse de croître. Dans ce cas, il n'y a pas de convergence. Pour une revue plus exhaustive sur la technique du maximum de vraisemblance, le lecteur pourra se diriger vers le mémoire de (Mai-Anh, 2012).

Dans cette partie, nous avons vu une méthode pour estimer les paramètres du modèle. À présent, nous allons aborder la phase de validation, pendant laquelle nous vérifions si les paramètres estimés $\hat{\beta}$ sont proches des paramètres du modèle de référence que nous noterons $\tilde{\beta}$.

2.4.3 Validation de l'estimateur du maximum de vraisemblance

Les techniques issues de l'analyse statistique nous permettront de faire de l'inférence statistique sur les paramètres. La procédure proposée est la suivante:

1. Faire une hypothèse (H_0), aussi appelée hypothèse nulle qui sera soit rejetée, soit non rejetée.
2. Spécifier l'hypothèse alternative (H_1).
3. Choisir le test statistique pour tester (H_0).
4. Déterminer un niveau de confiance et la taille de l'échantillon.
5. Trouver la distribution de la statistique sous l'hypothèse (H_0).
6. Énoncer la règle de décision.
7. Conclure sur le rejet ou non rejet de (H_0).

L'estimateur du maximum de vraisemblance possède des propriétés intéressantes asymptotiquement puisqu'il converge en probabilité pour chaque échantillon vers la valeur du paramètre de la population. De plus, sa loi de probabilité tend vers une loi normale et il est asymptotiquement efficace. Le test choisi est le test de Student avec un seuil de signification de 0,05 pour un échantillon d'observations de taille N . La statistique de Student $t = \frac{\hat{\beta}_q - \tilde{\beta}_q}{\hat{\mu}_q / \sqrt{N}} \forall q \in Q$, suit asymptotiquement une loi normale centrée réduite $N(0, 1)$. Sous H_0 , nous avons:

$$\Pr(L_{0,025} \leq t \leq L_{0,975}) = 0,95$$

c'est-à-dire, puisque t suit une loi normale centrée réduite asymptotiquement:

$$\Pr(-1,96 \leq t \leq 1,96) = 0,95$$

La règle de décision est donc: on rejette H_0 lorsque la valeur absolue de la statistique de Student est supérieure au 0,975 quantile de la loi $N(0, 1)$, $L_{0,975}$, c'est-à-dire lorsque $|t| \geq 1,96$ puisqu'il s'agit d'un test bilatéral. Nous souhaitons faire un test de conformité en testant (H_0) : $\hat{\beta}$ est significativement proche de $\tilde{\beta}$. On a alors H_1 : $\hat{\beta}$ est significativement différent de $\tilde{\beta}$. On va tester (H_0) avec un test de Student pour un seuil de signification de 0,05. La variable de Student t suit asymptotiquement une loi normale centrée réduite $N(0, 1)$. La règle de décision est: on rejette H_0 lorsque la valeur absolue de la statistique de Student est supérieure au 0,975 quantile de la loi $N(0, 1)$, c'est-à-dire lorsque $|t| \geq 1,96$. Nous verrons plus tard avec notre modèle que les valeurs des variables de Student sont inférieures en valeur absolue à 1,96. On ne rejette donc pas H_0 . La seule conclusion qu'on puisse faire est qu'on "accepte" (H_0),

ce qui n'est pas équivalent à dire que (H_0) est vraie et (H_1) est fausse. Seulement, il n'y a pas de preuve nette que (H_0) soit fausse. Il est admis de dire que d'après nos résultats, $\hat{\beta}$ est significativement proche de la vraie valeur du vecteur de paramètres $\tilde{\beta}$ du modèle de référence.

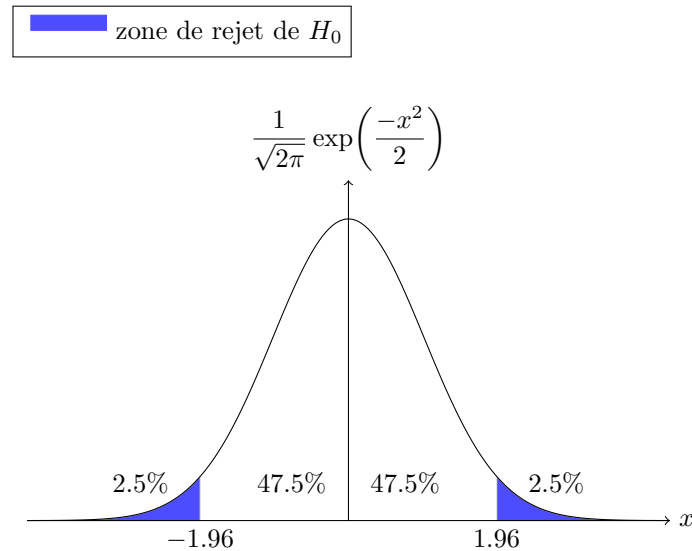


Figure 2.3 – Test de Student bilatéral à un seuil de 0,05 de signification

2.5 Des défis inhérents aux problèmes de choix de routes

2.5.1 Des problèmes combinatoires et de corrélation

Suivant la nature du problème à traiter, le respect des deux premières caractéristiques des ensembles d'alternatives (voir la section 1) peut présenter des défis plus ou moins difficiles à relever. Si on s'intéresse aux moyens de transport qu'un citoyen doit choisir pour se rendre à son restaurant préféré, on peut facilement définir l'ensemble des alternatives à {vélo, autobus, marche, métro}. Dans le cas des problèmes de choix de routes, particulièrement dans des régions développées et urbaines, on doit définir l'ensemble des alternatives en répondant d'abord à des questions comme les

suivantes: *Combien de routes existe-t-il entre une origine et une destination quelconques? Parmi ces routes, lesquelles constituent l'ensemble des alternatives?* La première question relève une difficulté inhérente aux réseaux que nous étudions car ils représentent des grandes zones urbaines et développées. Si nous considérons que des chemins peuvent contenir des boucles, alors le nombre de chemins possibles entre une origine et une destination quelconques tend vers l'infini, ce qui viole une des caractéristiques de l'ensemble des alternatives. Cette question de dénombrement sera aussi à se poser lorsque plus tard, nous chercherons à calculer le maximum de vraisemblance d'une observation partielle grâce à un sous-réseau, tel que nous verrons dans la section 4.2.2 avec l'équation (4.11).

Même si on se trouve dans une situation où les alternatives sont nombreuses mais en nombre fini, la deuxième question est difficile puisqu'en pratique, l'analyste ne connaît pas l'ensemble des routes considérées par le décideur. Ces problèmes ne touchent pas tous les modèles de choix de routes. Il faut préciser qu'il existe deux catégories de modèles qui se distinguent par la façon dont se construit leur ensemble d'alternatives. La première se base sur une construction dynamique arc par arc alors que la deuxième propose des routes directement, de manière statique. Ceux de la dernière catégorie sont précisément concernés par cette question. Pour estimer ces modèles, il est nécessaire d'avoir recours à un algorithme de génération de routes pour construire des ensembles d'alternatives, ce qui entraîne des calculs supplémentaires et un besoin accru en mémoire pour le stockage sans oublier que la route choisie par le décideur n'est pas garantie de se trouver dans l'ensemble généré. Des modèles tels que le Path Size Logit (PSL), le c-logit, puis le nested logit vu dans (Vovsha et Bekhor, 1998) ainsi que le mixed logit vu dans (Bekhor, Ben-Akiva et Ramming, 2002) et (Frejinger et Bierlaire, 2007) ont été estimés de cette façon. Les estimations des modèles se basant sur les ensembles d'alternatives générées seront biaisées, puisque les estimateurs varient avec ces échantillonnages de routes. Si on imagine un réseau routier comme un ensemble d'arcs et de nœuds représentant respectivement les portions de routes et les intersections, alors les routes sont constituées d'arcs et il est très probable que l'utilisation de l'algorithme de génération de routes aboutisse à des alternatives corrélées entre elles, c'est-à-dire lorsque les routes se chevauchent et partagent des arcs. Si on ne tient pas compte des corrélations, cela peut entraîner des risques d'obtenir des probabilités de choisir un chemin erronées. Nous verrons dans la section 2.6 comment on calcule ces probabilités. Il existe des

modèles incluant des corrections de corrélation remarquables par leur capacité à corriger avec une heuristique en gardant une structure simple, c'est le cas notamment du PSL dans (Ben-Akiva et Bierlaire, 1999) et du c-logit dans (Cascetta, Nuzzolo, Russo et Vitetta, 1996). Le modèle logit récursif propose aussi une correction de corrélations.

Nous venons de voir que les problèmes de choix de routes entraînent des problèmes de corrélation, ainsi que des complications de nature combinatoires par rapport à d'autres problèmes de choix discrets et qu'ils imposent des corrections de corrélation et des stratégies d'échantillonnage. La mise en œuvre de ces dernières impliquera que les paramètres des modèles peuvent ne pas converger vers la même valeur suivant l'échantillon obtenu. Nous continuerons néanmoins notre analyse de choix de routes avec un modèle de choix discrets, car il existe une autre approche de modélisation qui permet de s'affranchir de ces contraintes d'échantillonnage. Nous présenterons dans la section 2.6 le modèle logit récursif pour illustrer notre propos.

2.5.2 Les défis sur les observations

C'est avec des observations que nous pourrions estimer un modèle et leur récolte constitue par conséquent une étape critique. Dans cette section, nous présentons des techniques permettant de reconstituer arc par arc, avec une limite dans leurs précisions, des routes empruntées par des usagers. Ces techniques peuvent être appliquées par l'analyste qui veut étudier les choix de routes des usagers. Cependant, les observations disponibles ne sont pas forcément obtenues par ces techniques et souvent, l'analyste se sert de bases de données qui ont été construites pour des motifs variés et il devra en extraire les observations exploitables. La plupart du temps, ces dernières seront partielles et le principal défi sera d'être capable de s'en servir malgré tout pour faire l'estimation d'un modèle. Nous verrons ensuite comment exploiter des observations partielles.

2.5.3 Techniques de collecte des observations

Jusque là, nous avons introduit un cadre théorique pour formuler un modèle de choix de routes et nous avons vu une méthode pour faire des estimations. Dans la meilleure des situations, l'analyste a accès à toutes les données pertinentes et aux observations dont il a besoin.

La récolte des observations est une étape difficile et dans bien des cas, il faut se résoudre à travailler avec des bases de données incomplètes. Dans les problèmes de choix de routes, la connaissance des itinéraires avec les origines et les destinations est idéale mais elle reste difficile d'accès pour plusieurs raisons. La collecte d'observations consiste à reconstituer des chemins morceau par morceau et on distingue deux façons de procéder. La première façon consiste à interroger directement le décideur sur le trajet exact qu'il a suivi alors que l'autre manière de procéder propose de reconstruire le chemin indirectement par l'intermédiaire d'un GPS. Il existe dans la littérature des études pour illustrer la première méthode, par exemple dans (Mahmassani, Joseph et Jou, 1993) et dans (Abdel-Aty, Kitamura, Jovanis, Reddy et Vaughin, 1995). Dans (Prato, 2004), l'auteur exploite des technologies du web pour demander aux participants d'indiquer sur une carte interactive leur route ainsi que des routes alternatives. Avec 236 participants, 575 routes ont été définies dont 276 étaient finalement retenues pour l'étude car le reste était composé de routes incomplètes. On peut se demander cependant dans quelle mesure une personne est capable de se souvenir du trajet emprunté. Pour notre étude, les seules observations qu'on puisse espérer obtenir sont les arcs sur lesquels des décideurs se sont fait contrôler. Au mieux nous disposerions d'observations partielles et nous n'aurions pas le luxe de pouvoir choisir entre des observations complètes ou partielles. La deuxième approche pour reconstituer une route utilise la technologie GPS et permet de capturer de grandes quantités de données sur les positions ainsi que les heures associées. On peut connaître un temps de départ et d'arrivée par exemple.

Lorsqu'un analyste dispose d'observations qui sont par exemple des routes définies sans ambiguïtés, en incluant l'origine, la destination et les portions de route exactement utilisées, son travail pour estimer les paramètres du modèle est facilité. En pratique, on a vu que l'obtention des observations est une étape difficile et il faut se contenter d'observations partielles. Utiliser ces observations demande des précautions parce que l'analyste peut être tenté de faire des traitements additionnels et il est raisonnable de penser que moins il y a de traitements sur les observations moins les estimations du modèles seront biaisées. Dans le cas où des observations sont obtenues de manière passive avec un GPS, il se peut qu'elles ne correspondent pas au réseau qui a été modélisé. Nous décrivons dans ce qui suit une méthode qui permet d'utiliser ces observations sans faire de traitements additionnels. Plus précisément, cette méthode permettra de faire coïncider des observations brutes issues du GPS

avec des éléments du réseau modélisé.

2.5.4 Traitement des observations partielles

Ce passage reprend l'étude de (Bierlaire et Frejinger, 2008). Il arrive que l'analyste soit obligé de manipuler des observations brutes pour obtenir un ensemble d'observations exploitable. C'est le cas lorsque celles-ci sont partielles et l'analyste peut y introduire des arcs manquants en privilégiant arbitrairement, par exemple, ceux qui procurent une route plus courte. Dans (Bierlaire et Frejinger, 2008), les auteurs montrent qu'un analyste peut manipuler des observations sans introduire de biais ni d'erreurs, ceci grâce à un élément de modélisation, le domaine de pertinence d'une donnée (DPD). Le DPD définit une aire dans laquelle la donnée est pertinente. Par exemple, une donnée correspondant à un point GPS ne figure pas forcément sur le réseau routier. Ce point a des chances de se retrouver seulement dans un voisinage d'un arc ou d'un sommet du graphe défini pour la modélisation de ce réseau. Le DPD peut alors se matérialiser par exemple par un disque dont le centre est le point GPS et le rayon est tel que le disque comprend au moins un élément du graphe modélisant le réseau. L'observation i , liée au décideur n , est une séquence (d_1, d_2, \dots, d_k) d'observations brutes (points GPS, relevé d'arcs, etc ...) et le défi pour l'analyste est de retrouver la probabilité $\Pr_n(i)$ que le décideur n choisisse i , sachant l'ensemble des couples origine-destination S_i . S étant l'ensemble des couples origine-destination du réseau modélisé et $\delta(d, e) = 1$ si le DPD associé à la donnée brute d peut être associée à une donnée e du réseau modélisé, nous obtenons:

$$S_i = \{s \in S \mid \delta(d_1, s_0)\delta(d_k, s_d) = 1\} \quad (2.24)$$

En définissant les DPD de telle sorte qu'au moins un couple (s_0, s_d) satisfasse (2.24) et que S_i soit non vide, nous déduisons:

$$\Pr_n(i) = \sum_{s \in S_i} \Pr_n(s|S_i) \sum_{\sigma \in J_n(s)} \Pr_n(i|\sigma) \Pr_n(\sigma|J_n(s); \beta) \quad (2.25)$$

$J_n(s)$ étant l'ensemble des chemins en fonction de s parmi lesquels n choisit un chemin σ .

$\Pr_n(s|S_i)$ est la probabilité que le couple origine-destination courant est s sachant S_i .

$\Pr_n(i|\sigma)$ est la probabilité d'observer i si le chemin choisi par n est σ .

$\Pr_n(\sigma|J_n(s); \beta)$ est le modèle de choix de route, β étant un vecteur de paramètres à déterminer.

Le DPD fait correspondre à l'observation $i = (d_1, d_2, \dots, d_k)$ un chemin $\sigma = (l_1, l_2, \dots, l_p)$, sans avoir besoin de manipuler i . Une même observation peut être associée à plusieurs chemins suivant la définition du DPD. Cette méthodologie nous permettra plus tard de travailler avec des chemins complètement définis à partir d'observations incomplètes basées sur des reports de portions de routes. Lorsque l'on parle d'associer une observation à un ou plusieurs chemins sur le réseau modélisé, on peut anticiper une difficulté de nature combinatoire. La méthodologie que nous venons de voir sera associée à un modèle appelé récursif logit. Nous présentons en détails dans la section suivante ce modèle qui nous servira à étudier les choix de routes des camionneurs. Il présente quelques avantages qui expliquent pourquoi nous l'avons précisément choisi.

2.6 Modèle logit récursif

2.6.1 Un choix de routes dynamique

Cette partie reprend les informations vues dans (Fosgerau, Frejinger et Karlström, 2013). Nous avons choisi ce modèle en raison des atouts qu'il présente. La nature dynamique du processus de choix de routes a pour conséquence que nous n'avons plus besoin de constituer des échantillons pour définir nos ensembles d'alternatives. En donnant la possibilité de calculer les flux dans un réseau, il peut servir d'outil pour des gestionnaires de réseaux routiers sans compter que l'estimation de ses paramètres est simplifiée grâce à la formulation logit de la probabilité de choix de routes. Nous savons que le problème des corrélations de routes peut entraîner des erreurs dans les calculs de probabilité mais le logit récursif offre une solution, bien que coûteuse, pour les limiter. Modélisons le réseau routier par un graphe orienté $G(\nu, A)$ où A est l'ensemble des arcs, ν l'ensemble des sommets. On notera que a et k représentent des arcs, et les arcs accessibles depuis k sont dans l'ensemble $A(k) \subset A$. La maximisation de l'utilité sera traitée comme un problème de programmation dynamique. On dira qu'au moment du choix le décideur est dans l'état k , alors que a tel que $a \in A(k)$ est une action potentielle sachant l'état k .

Le décideur n étant rationnel, il choisit a de telle sorte que l'utilité qu'il en retire est maximale. Celle-ci est la somme de l'utilité déterministe $v_n(a|k)$ avec l'espérance du maximum de l'utilité de a vers la destination $V_n^d(a)$ et du terme d'erreur de type I et s'écrit donc ainsi:

$$u_n(a|k) = v_n(a|k) + V_n^d(a) + \mu\epsilon_n(a) \quad \forall k \in A. \quad (2.26)$$

d est un arc ajouté à $G(\nu, A)$ désignant la destination puis représentant un état absorbant. Nous avons $v_n(d|k) = 0$, pour tout arc k ayant comme destination l'arc d . Sinon, $v_n(a|k) < 0$ pour tout arc dans $A(k)$.

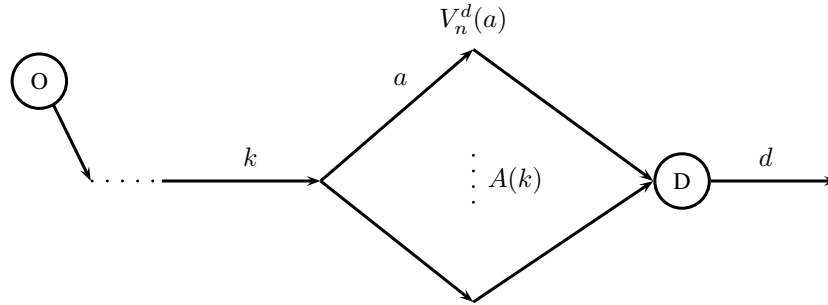


Figure 2.4 – Illustration de la notation

2.6.2 Un modèle de choix de routes sans échantillonnages sur l'ensemble des alternatives

L'espérance du maximum d'utilité $V_n^d(k)$ s'obtient en résolvant l'équation de (Bellman, 1957):

$$V_n^d(k) = E[\max_{a \in A(k)} (v_n(a|k) + V_n^d(a) + \mu\epsilon_n(a))], \quad \forall k \in A \quad (2.27)$$

ce qui peut aussi s'écrire:

$$\begin{aligned} V_n^d(k) &= Eu_{\max} \\ &= \mu \ln \sum_{a \in A(k)} e^{\frac{1}{\mu}(v_n(a|k) + V_n^d(a))} \quad \forall k \in A. \end{aligned}$$

et en introduisant le symbole $\delta(a|k)$ tel que:

$$\delta(a|k) = \begin{cases} 1 & \text{si } a \in A(k) \\ 0 & \text{sinon} \end{cases} \quad (2.28)$$

Nous avons:

$$V_n^d(k) = \begin{cases} \mu \ln \sum_{a \in A} \delta(a|k) e^{\frac{1}{\mu}(v_n(a|k) + V_n^d(a))} & \forall k \in A. \\ 0 & k = d \end{cases} \quad (2.29)$$

Il a été démontré dans (McFadden, 1978) et (Fosgerau, Frejinger et Karlström, 2013) que la probabilité qu'un décideur n choisisse l'alternative j peut s'obtenir en prenant le gradient de Eu_{\max} par rapport à la variable $v_n(a|k) + V_n^d(a)$.

$$\Pr_n^d(a|k) = \frac{e^{\frac{1}{\mu}(v_n(a|k) + V_n^d(a))}}{\sum_{a' \in A(k)} e^{\frac{1}{\mu}(v_n(a'|k) + V_n^d(a'))}} \quad (2.30)$$

On remarque que le dénominateur est aussi $e^{\frac{1}{\mu}V(k)}$. D'après les propriétés markoviennes du modèle et en sachant qu'une route σ est définie par une suite d'arcs $\{k_i\}_{i=0}^I$, où k_0 est l'origine et k_I est la destination, on en déduit, à l'aide de (2.30) la probabilité que le preneur de décisions n passe par le chemin σ est :

$$\Pr_n(\sigma) = \prod_{i=0}^{I-1} \Pr_n(k_{i+1}|k_i) \quad (2.31)$$

$$= \prod_{i=0}^{I-1} e^{\frac{1}{\mu}(v(k_{i+1}|k_i) + V(k_{i+1}) - V(k_i))} \quad (2.32)$$

$$= e^{-\frac{1}{\mu}V(k_0)} \prod_{i=0}^{I-1} e^{\frac{1}{\mu}(v(k_{i+1}|k_i))} \quad (2.33)$$

En notant $v(\sigma) = \sum_{i=0}^{I-1} v(k_{i+1}|k_i)$, nous avons:

$$\Pr_n(\sigma) = \frac{e^{\frac{1}{\mu}(v(\sigma))}}{e^{\frac{1}{\mu}V(k_0)}} = \frac{e^{\frac{1}{\mu}v(\sigma)}}{\sum_{\sigma' \in \Omega} e^{\frac{1}{\mu}v(\sigma')}} \quad (2.34)$$

Où Ω est l'ensemble de tous les chemins possibles du réseau, dont le cardinal tend vers l'infini. Cette expression de la probabilité de choisir un chemin est quasiment

identique à celle qui a été dérivée pour le logit en (2.11), mais avec Ω remplaçant J_n . L'avantage de spécifier des chemins dynamiquement est qu'il n'est plus besoin de se préoccuper d'une stratégie d'échantillonnage, évitant ainsi à la fois une source de biais et de calculs.

En considérant que nous avons un problème à horizon non infini de facteur 1 et que l'environnement est déterministe puisqu'un état k et une action a mènent à un état déterminé, l'équation (2.29) devient:

$$e^{\frac{1}{\mu}V(k)} = \begin{cases} \sum_{a \in A} \delta(a|k) e^{\frac{1}{\mu}(v_n(a|k) + V_n^d(a))} & \forall k \in A. \\ 1 & k = d \end{cases} \quad (2.35)$$

On peut écrire sous forme matricielle (2.35) en fonction de M , définie plus bas, qui est la matrice d'incidence relative aux utilités instantanées.

$$M_{ka} = \begin{cases} \delta(a|k) e^{\frac{1}{\mu}v_n(a|k)} & \forall a \in A(k). \\ 1 & \text{sinon} \end{cases} \quad (2.36)$$

Puisque d est un arc artificiel et qu'il n'appartient pas à G , il n'a pas de successeur et la ligne correspondant à d est un vecteur nul. On notera $\tilde{A} = A \cup d$. Soient φ et b des vecteurs de dimensions $|\tilde{A}|$ tels que $\varphi = e^{\frac{1}{\mu}V(k)}$, $b_k = 0$, $k \neq d$ et $b_d = 1$. L'équation (2.36) s'écrit alors:

$$\varphi = M\varphi + b \Leftrightarrow (I - M)\varphi = b. \quad (2.37)$$

Le système (2.37) admet une solution si $I - M$ est inversible, ce qui est le cas lorsque M^m tend vers zéro quand m tend vers ∞ en étant un entier. M^m étant la matrice d'incidence relative aux utilités lorsque le décideur prend m actions depuis son état d'origine, elle contient les exponentielles des utilités pour des chemins de longueur m entre n'importe quelle paire d'arcs.

Si on suppose que $v(a|k)$ est linéaire en β alors $\frac{1}{\mu}v(a|k) = \frac{1}{\mu}\beta^T y$, y étant un vecteur de variables exogènes observables par l'analyste et β est un vecteur de dimension $Q \geq 1$. L'inversibilité de $I - M$ dépend de la valeur des quotients $\frac{\beta_q}{\mu}$, en considérant y constant. Il se peut parfois que $\frac{\beta_q}{\mu}$ soit fortement négatif; dans ce cas, M^m tend vers zéro quand m tend vers l'infini et $I - M$ est inversible. Mais, chaque $\frac{\beta_q}{\mu}$ dépend du réseau, car il varie avec le facteur d'échelle μ . Un vecteur β peut entraîner que

$I - M$ soit inversible dans un réseau mais pas dans un autre. Plus tard, lorsque les valeurs de β devront être estimées, le mieux est de donner un vecteur initial dont les composantes ont des valeurs absolues élevées au logiciel d'estimation pour éviter que $I - M$ soit non inversible. Au préalable, nous fixerons la valeur de μ à 1.

2.6.3 Prédiction de flux dans un réseau

Nous venons de voir que le modèle récursif logit permet de surmonter le problème de l'échantillonnage. Un autre de ses atouts est qu'il peut servir pour la prédiction des flux dans un réseau. Même si ce n'est pas un aspect qui sera développé dans ce projet, il est utile de le mentionner, car nous gardons en tête le côté applicatif qui peut offrir une perspective intéressante pour les gestionnaires de réseaux routiers par exemple. Des logiciels permettant d'estimer des flux de véhicules existent déjà sur le marché mais il est permis de croire que les modèles de choix discret de routes peuvent apporter une autre contribution. Ces calculs de prédiction se font avec l'aide de la relation (2.30) qui donne la probabilité de choisir un arc suivant son état courant. Puisque nous avons introduit un arc fictif d , nous pouvons commencer par calculer des flux sur des chemins partant de diverses origines vers d . Ensuite, il suffira de changer de destination et le flux global dans le réseau pourra être calculé en superposant tous les flux calculés pour chaque destination. Ce qui fait que, pour chaque arc a , le flux espéré $F(a)$ est la somme du flux engendré en a finissant en d avec celui arrivant en a depuis les arcs k . On a alors $F(a) = G(a) + \sum_{k \in A} \text{Pr}(a|k)F(k)$. En posant $F = (F(k)_{k \in A})^T$ et $G = (G(k)_{k \in A})^T$, le flux en a s'écrit:

$$(I - P^T)F = G. \quad (2.38)$$

La résolution de ce système nécessite que $I - P^T$ soit inversible, ce qui est démontré dans (Baillon et Cominetti, 2008).

2.6.4 Un correctif pour la corrélation

Non seulement les prédictions de flux ont un intérêt pratique mais aussi, ils peuvent servir dans la correction des problèmes de corrélation. En effet, ils concernent aussi le modèle logit récursif mais contrairement aux méthodes développées dans la section 2.5.1, celle employée dans ce modèle ne demande pas d'énumérations de

chemins au préalable. Le fait d'utiliser les flux pour un correctif est pertinent car ils sont un des attributs de chaque arc et le modèle se base sur les arcs pour construire un chemin. Pour un chemin non corrélé, nous avons $F(a) = G(a) + \Pr(a|k)F(k)$ tel que k est l'unique prédécesseur de a alors que s'il y a de la corrélation, nous avons $F'(a) = G(a) + \sum_{k \in A} \Pr(a|k)F(k)$. Par conséquent, il y a de la corrélation lorsque $F'(a) > F(a)$. Le but d'un correctif est de tenir compte du phénomène de corrélation, ce qui est systématique lorsqu'on traite de choix de routes, afin de réduire des erreurs dans les calculs de probabilité. Le Link Size (LS) vu dans (Fosgerau, Frejinger et Karlström, 2013) a l'avantage d'être compatible avec notre modèle, mais il dépend de la spécification de la fonction d'utilité et il reste à analyser l'impact dans l'estimation des paramètres par rapport aux différentes valeurs du LS. L'introduction du LS implique de connaître les flux sur le réseau, ce qui entraîne des calculs multi-origine, multi-destination et par conséquent, l'estimation du modèle est rendue plus difficile. En résumé, notre choix s'est porté sur le récursif logit, qui par nature nous évite de travailler avec des ensembles d'alternatives constitués d'échantillons de routes. La possibilité de prédire des flux sera appréciée pour les personnes en charge du management d'un réseau routier. Le calcul des probabilités fait intervenir une fonction logit et nous aurons des simplifications bienvenues lorsque nous passerons à l'étape d'estimation du modèle. Enfin, il permet d'intégrer un correctif pour traiter le problème de corrélation de routes. Dans ce qui va suivre, nous allons montrer comment expliquer l'utilité en fonction des données à notre disposition et nous pourrons en déduire la probabilité qu'un camionneur choisisse sa route.

Chapitre 3

Spécification du modèle

Nous allons dans ce chapitre reprendre les résultats sur le modèle logit récursif et les adapter au contexte de notre problème de choix de routes des camionneurs sur le réseau routier québécois. Pour spécifier notre modèle, il faut au préalable se poser la question suivante: *Quels peuvent être les attributs explicatifs pour la fonction d'utilité?* Ensuite, nous devons collecter des données liées à ces attributs. L'analyste doit déployer des moyens pour collecter les données qu'il a ciblées ou bien, comme dans notre cas, il faut se baser sur des données qui ont été enregistrées dans d'autres buts que celui de notre étude. Nous nous sommes servis des informations qui se trouvent dans le rapport de ([Gendron, Crainic, Florian, Metnani et Deslauriers-Gaboury, 2013](#)) puis nous avons identifié et classé les attributs collectés sous forme de tableaux, permettant ainsi à de futurs analystes d'avoir une idée claire de ce dont ils disposent. Ce travail de description de données a été un travail intéressant au début de ce projet de mémoire, car il a occasionné des interactions avec des chercheurs et du personnel de CRQ, ce qui constitue une expérience stimulante. Nous verrons que les données peuvent être soit des données représentant des attributs du réseau routier, soit des données liées à un contrôle. Dans le premier cas, elles seront très utiles pour la modélisation du réseau routier et dans le dernier cas, ces données vont constituer une observation dont nous donnerons une définition puis, nous expliciterons la fonction permettant de calculer la valeur de l'utilité instantanée. Ce chapitre se termine par une formulation permettant de calculer une probabilité de choix d'alternatives par un décideur de manière dynamique.

3.1 Récolte de données et notion d'observation

3.1.1 La récolte et la description de données

Nous ne disposons pas seulement de données sur le réseau de transport. En effet, grâce aux renseignements fournis par le CRQ, nous pouvons compléter notre connaissance de la situation avec des attributs observables concernant à la fois les alternatives et les décideurs. Les données de CRQ n'avaient pas pour finalité d'être utilisées pour une étude de comportement des usagers sur le réseau du Québec. Pour l'analyste, il s'agit juste de saisir l'opportunité de travailler avec des données existantes et pouvant contenir des informations pertinentes. C'est pourquoi certaines données font cruellement défaut notamment celles concernant les origines, les destinations et le trajet total, mais aussi celles concernant les perceptions du temps et de la qualité du trajet et pour finir, celles qui sont liées aux attitudes.

De jour comme de nuit, les patrouilles de CRQ sont déployées, alors que des sites fixes de contrôles sont éparpillés sur une grande partie du territoire québécois. Les mouvements de transport vérifiés (MTV) enregistrent des informations sur la nature des infractions, sur les catégories de transporteurs et peuvent indiquer le lieu de l'interception dans certains cas. La quantité de données saisie lors d'un MTV peut varier suivant la situation puisque sur l'ensemble des informations que peut recueillir un agent de contrôle, certaines sont optionnelles. Par exemple, lorsqu'une infraction a toutes les chances d'aboutir à une action judiciaire, le nom de la municipalité dans laquelle elle s'est produite est enregistré. Aussi, des données socio-économiques sur le conducteur peuvent aider à affiner notre étude sur leur comportement. Il peut s'agir de leur âge, de leur genre, de leur nationalité, ainsi que de toute autre information figurant dans leur permis de conduire. Nous connaissons les différents types de routes, comprenant celles qui sont interdites aux camionneurs et que certains empruntent malgré tout. Des routes peuvent être dans des zones accidentogènes, et par habitude, les camionneurs les connaissent et peuvent décider de les contourner ou bien de les prendre si le détour prend trop de temps. Il va sans dire que ces comportements à risque peuvent occasionner des troubles à la sécurité publique. Bien qu'il soit pertinent de faire des hypothèses sur l'influence de l'heure pour étudier des effets de saisonnalité, elles ne pourront pas être testées puisque de telles données ne sont pas encodées. Lorsque des agents de CRQ repèrent un véhicule qui semble ne pas être en

règle, ils procèdent à une interpellation puis, si les conditions de sécurité sont réunies, procèdent au contrôle, sinon ils doivent se déplacer vers une aire sécurisée. En effet, une des non conformités souvent répertoriées est l'excès de poids du véhicule et les agents doivent pouvoir en effectuer une mesure grâce à des opérations qui nécessitent des manipulations délicates. Le lieu d'interception se trouve sur une portion de route et dans la plupart du temps, il faut amener le véhicule faisant l'objet du contrôle dans un endroit adéquat. Le MTV permettra d'enregistrer des informations sur l'infraction, telles que le type d'infraction, le montant de l'amende, ainsi que le lieu où s'est effectué le contrôle. Pour l'analyste, cette dernière information est une source d'erreur puisqu'il étudie le choix de routes du décideur. Or, le fait de se retrouver sur un arc sécurisé est indépendant de sa volonté et ce genre de situation où la position du décideur est inexploitable arrive vraisemblablement lorsqu'il s'agit de contrôles effectués par des patrouilles mobiles. Au contraire, les données provenant des MTV effectués dans des sites de type infrastructure seront plus fiables puisque c'est le décideur qui choisit de rouler sur la portion de route contrôlée. Cette dernière remarque nous indique que les échantillons d'observations ne seront pas aléatoires et qu'un biais de sélection est à anticiper. Nous avons un aperçu des attributs de réseau accessibles que nous exploiterons dans le tableau I. Dans le tableau II, les attributs sont pertinents, mais ils ne sont pas encore disponibles. On peut penser que pour la variable $x_{accident}(a)$, associée à un arc a , sa valeur est par défaut 0 et qu'elle passe à 1 lorsque le nombre d'accidents dépasse un seuil. La variable $x_{contour}(a)$ prendrait la valeur 1 lorsque l'arc a a comme successeur un arc d'évitement. La valeur de la variable $x_{intérêt}(a)$ prendrait la valeur 1 si l'arc a mène à une zone industrielle par exemple. Les données du tableau III sont aussi en attente, certaines informations comme l'âge et le type de véhicule peuvent être intéressantes pour des segmentations futures de la population. D'autres informations sont disponibles mais nous ne pouvons pas en tenir compte étant donné que ce sont des données calculées comme les nombres annuels moyens de véhicules sur un arc. Nous n'avons aucun contrôle sur la façon dont les calculs et les manipulations ont été faits, c'est pourquoi nous ne montrons que des données brutes.

Nous venons de donner une description des données liées au réseau et provenant de MTVs. Ces dernières constituent les observations qui vont nous être essentielles pour la phase d'estimation des paramètres de notre modèle.

Variable	Description	Valeurs
$G(A,\nu)$	Ensemble des arcs du réseau du Québec avec attributs liés aux arcs (distance, vitesse maximale, catégorie de route)	$ A =10320$, $ \nu =4946$ pour la province de Québec. $ A =415$, $ \nu =224$ pour Beloeil
A^S	Ensemble des arcs pouvant accueillir un site d'intervention, qu'il soit un site d'infrastructure ou un site sécurisé servant aux interventions des patrouilles	
A^I	Ensemble des arcs de A^S ne pouvant accueillir qu'un site d'infrastructure	
A^P	Ensemble des arcs de A^S ne pouvant accueillir qu'un site sécuritaire	

Table I – Attributs du réseaux

3.1.2 Les observations

Un transporteur n devra au cours de sa route σ_n immobiliser son véhicule s'il est interpellé par des agents mobiles de la CRQ ou lorsqu'il entre dans un site de contrôle de type infrastructure. Les lieux d'interception, les données liées au conducteur et les autres attributs du MTV enregistrés pour n , quand ils sont connus, peuvent composer une observation. Comme nous l'avons mentionné dans le chapitre 1, nous utiliserons des observations synthétiques, car nous serons capables de conclure si nos estimations sont valides. Par souci de simplicité, nous décidons de ne retenir que les lieux d'interception pour simuler une telle observation. En nous basant sur la théorie des graphes, nous disons que les portions de route et les intersections sont modélisées par des arcs et des nœuds respectivement. Les lieux d'interception se trouvent soit sur des arcs soit sur un nœud. Nous traitons les cas où ce sont des arcs car en pratique, non seulement les interceptions se font très rarement sur les croisements, mais en plus ces derniers ne sont qu'exceptionnellement encodés dans un MTV lorsque le cas se présente. Notre connaissance de la route prise par le décideur est d'autant plus fine que le nombre d'interceptions est élevé. Dans un cas idéal, il y aurait un système GPS sur chaque camion ou bien on peut penser à installer des postes de contrôle sur exactement chaque arc et on observerait $\{k_i\}$ tel que $k_{i+1} \in A(k_i)$, $i = 0 \dots I - 1$. σ_n est la suite de tous les arcs empruntés par n et modélise

Variable (non disponible)	Description	Valeurs
$x_{accident}(a)$	Paramètres indiquant si l'arc appartient à une route accidentogène	0 ou 1
$x_{contour}(a)$	Paramètres indiquant si l'arc appartient à une route de contournement	0 ou 1
$x_{intérêt}(a)$	Paramètres indiquant un site d'intérêt avoisinant	0 ou 1
$N(a)$	L'ensemble des arcs sur lesquels tout véhicule intercepté par une patrouille sur un arc a pourra être contrôlé de manière sécurisée	

Table II – Attributs du réseaux

ainsi la route du décideur. En considérant que k_0 et k_I sont l'origine et la destination respectivement, nous avons: $\sigma_n = \{k_i\}$, $i = 0 \dots I$. Les ressources de contrôle étant limitées, nous observerons plutôt σ'_n , une sous-suite de σ_n , $\sigma'_n = \{k'_i\}$, $i = 0 \dots I$. k'_0 et k'_I ne sont pas forcément les arcs désignant l'origine et la destination et nous n'avons plus nécessairement $k'_{i+1} \in A(k'_i)$, $i = 0 \dots I - 1$. Ceci nous conduit donc à simuler dans un premier temps des observations complètes puis, afin de simuler le manque d'informations, nous enlèverons dans chaque observation obtenue précédemment des arcs et nous obtiendrons des ensembles d'observations partielles. Pour créer ces observations, nous devons d'abord spécifier une fonction régissant le comportement des camionneurs.

3.2 Modèle

Dans cette partie, nous commençons par présenter la phase de modélisation. Ensuite, nous spécifierons la fonction d'utilité instantanée et nous serons finalement en mesure de calculer les probabilités de choix d'alternatives de manière dynamique.

3.2.1 Modélisations du réseau et du comportement

Les données fournies par le MTQ nous renseignent sur les infrastructures, notamment la nature des routes qui peuvent être locales ou des autoroutes entre autres. La connaissance des longueurs ainsi que des vitesses maximales de chaque portion

Variable (non disponible)	Description	Valeurs
Id_{MTV}	Identification MTV	Base de données fournies par le CRQ
Mu	Municipalité	Nom de la municipalité
Id_{route}	Optionnellement, le numéro de la route	Un numéro de route
$x_{nord}, x_{sud}, x_{est}, x_{ouest}$	Paramètres indiquant le sens de déplacement	0 ou 1
$\nu_i \in \nu$	Une intersection, qui est un nœud du réseau	Numéro de nœud
$Id_{infraction}$	Type infraction	Code de l'infraction
$\hat{age}, genre, tarif$	Informations sur le conducteur (H/F) de camion, tarif amendes	Valeurs relevées
K	Ensemble de catégories de véhicules	Poids lourd ou léger, conforme ou non conforme, inspecté ou non inspecté

Table III – Attributs provenant d'un MTV

de route nous permettra de calculer leur temps minimal de parcours par un conducteur. Chaque portion de route est caractérisée par le type de site de contrôle qui y est présent lorsqu'il y en a un. Les décideurs, qui sont des transporteurs de marchandises, se déplacent sur le réseau routier du Québec. Celui-ci sera modélisé par un graphe orienté $G(\nu, A)$. Les portions de route entre chaque intersection seront modélisées par des arcs $k \in A$ et les intersections seront représentées par les sommets. Pour représenter le graphe, on peut utiliser une matrice d'incidence W_{inc} où chaque ligne est un vecteur qui indique les successeurs de k .

$$W_{inc}(k, a) = \begin{cases} 1 & \text{si } a \in A(k) \forall k \in A \\ 0 & \text{sinon} \end{cases} \quad (3.1)$$

Comme nous l'avons vu dans la revue de littérature dans la section 2.6.1, une alternative est un arc $a \in A(k)$, successeur de l'arc k sur lequel se situe le décideur. Cet ensemble d'alternatives est non vide car le décideur doit pouvoir se rendre depuis n'importe quelle origine jusqu'à n'importe quelle destination, et par conséquent, on supposera que le graphe $G(\nu, A)$ est fortement connexe. Le modèle ne prend pas en compte le niveau de congestion et suppose que chaque décideur se déplace à la

vitesse maximale autorisée tout au long de sa route comme s'il était seul. Il sera aussi supposé que chaque décideur connaît les points de contrôle fixes et que les localisations des sites de contrôle mobiles sont des informations que les décideurs se transmettent entre-eux, on considère qu'ils ont ces informations en quelques minutes. Si un décideur n emprunte un arc sur lequel il y a un site fixe de contrôle ou une patrouille, alors il se fera inévitablement contrôler. Pour vérifier la forte connexité du graphe modélisant le réseau, on pourra s'aider de l'algorithme de Tarjan, décrit dans (Tarjan, 1972). Les décideurs prennent leur décision indépendamment les uns des autres à chaque fois qu'ils se trouvent à une intersection. Comme ce qui a été vu précédemment, un décideur est dans l'état k et choisit l'action a tel que l'espérance de la somme de son utilité instantanée, $u_n(a)$ avec l'espérance du maximum d'utilité de a vers la destination d , $V_n^d(a)$, soit maximale.

Choisir de manière dynamique ses arcs depuis son état d'origine jusqu'à son état final revient à choisir une route pour le décideur. Pour pouvoir traiter ce problème de choix de routes dynamique comme un processus de décisions Markovien, il a fallu au préalable ajouter un arc artificiel d de destination qui correspond à un état absorbant. Nous pouvons définir la fonction d'utilité instantanée comme étant la somme d'une fonction d'utilité déterministe et d'un terme d'erreur.

$$u_n(a|k) = v_n(a) + \mu \epsilon_n(a) \quad \forall a \in A(k) \quad (3.2)$$

Le décideur va chercher à maximiser une utilité qui est la somme de l'utilité instantanée et de l'espérance du maximum d'utilité de a vers la destination d .

3.2.2 Spécification de la fonction d'utilité déterministe

La fonction d'utilité dépend des attributs qui caractérisent les différentes alternatives qui se présentent au décideur et des attributs liés au décideur lui-même. En tant qu'analyste, ces données pertinentes nous échappent et il nous est seulement possible de faire des hypothèses sur la forme de la fonction d'utilité. Celles-ci aboutissent à la définition de la fonction d'utilité déterministe v_n qui dépend d'attributs que nous pouvons observer. Deux questions importantes sont soulevées à ce niveau: ainsi, on doit, dans un premier temps, se demander quels éléments du vecteur des attributs observables (x_n, s_n) sont pertinents. Les données et les tests statistiques nous l'indiqueront. Ensuite, il faut réfléchir sur la nature de la fonction d'utilité. Avec une fonction logit,

nous supposons que la fonction d'utilité est linéairement dépendante du vecteur de paramètres β . De cette manière, la fonction de log-vraisemblance est concave, c'est pourquoi cette hypothèse est souvent utilisée d'après (Ben-Akiva et Lerman, 1985) au chapitre 5. Il ne s'agit pas d'une linéarité par rapport aux attributs.

À travers les informations fournies par le ministère des transports du Québec, on peut se focaliser sur la longueur des portions de route, la vitesse maximale y étant autorisée, ce qui nous permet d'en déduire le temps minimal $tt(a)$ pour parcourir un arc a . Les informations sur les lieux de contrôle nous sont fournies par le Contrôle Routier du Québec. Nous suggérons dans un premier temps une fonction v_n qui dépendra des attributs suivants, $tt(a)$, le temps minimal de parcours sur un arc et de $cr(a)$, qui indiquera la présence d'un contrôle sur l'arc. Puis, pour différencier deux routes dont les temps de trajets minimum sont égaux ainsi que le nombre de contrôles subis, on peut prendre en compte une variable qui compte le nombre de sections de routes empruntées $l(a)$. On aura $v_n = v_n(tt, cr, l, \beta)$, avec $\beta = (\beta_{tt}, \beta_{cr}, \beta_l)^T$ le vecteur des coefficients associés aux attributs du modèle. Les valeurs tt sont exprimées en minutes, cr est une variable binaire et vaut 0 lorsque l'arc est dépourvu de contrôles et vaut 1 sinon. La variable l vaut 1 puisqu'elle sert à compter les arcs choisis pour le chemin.

À la destination d , lorsque n est dans l'état d , $V_n(d) = 0$. Sinon, les utilités déterministes de n sont strictement négatives pour que l'équation de Bellman possède des solutions. Les trois attributs mentionnés plus tôt représentent des "coûts" pour le décideur, car à mesure que le temps de parcours minimal augmente, son utilité diminue. De même, le fait de prendre un arc ayant un point fixe ou mobile de contrôle aura pour conséquence de faire diminuer son utilité puisqu'il y a un risque que le contrôle révèle une infraction, sans compter la perte de temps, sachant que la durée d'une opération de contrôle varie en fonction de la nature des mesures à effectuer, telle que la prise de poids qui nécessite un appareillage particulier. Enfin, on suppose, à temps de parcours égal et avec un nombre de contrôle égal, qu'une route la plus directe possible, c'est-à-dire avec un changement minimum d'arcs est préférée par le décideur, d'où l'intégration de l'attribut l . La fonction d'utilité (3.2) s'écrit:

$$u_n(a|k) = \beta_{tt}tt(a) + \beta_{cr}cr(a) + \beta_l l(a) + \mu_{\epsilon_n}(a), \forall a \in A(k) \quad (3.3)$$

Nous faisons l'hypothèse que le vecteur de paramètres est le même pour toute la

population par souci de simplicité. Éventuellement, nous pourrions supposer et tester que la population est divisible en sous-groupes homogènes, mais ceci nécessiterait de connaître les variations de goût au niveau individuel par exemple. Pour marquer les différences de goût, on peut soit intégrer des attributs de chaque décideur dans la formulation de la fonction d'utilité déterministe ou bien il faudrait utiliser un modèle dans lequel on définit le vecteur de paramètres β comme une variable aléatoire et chaque β_n est indépendamment et log-normalement distribué.

Les trois attributs ont des valeurs non négatives et de ce fait, les composantes du vecteur β sont strictement négatives. Plus leurs valeurs absolues augmentent, plus les utilités déterministes diminueront. En plus de s'intéresser aux valeurs absolues, il est pertinent de regarder les ratios entre les différentes composantes de β . Prenons β_{cr} et β_{tt} et regardons le ratio $\frac{\beta_{cr}}{\beta_{tt}}$. Si la fonction d'utilité déterministe s'écrit:

$$v_n(a) = -4tt(a) - 8cr(a) - 0,2l(a), \forall a \in A(k) \quad (3.4)$$

Le ratio $\frac{\beta_{cr}}{\beta_{tt}}$ vaut $\frac{-8}{-4} = 2$ signifie que le décideur est prêt à prendre un chemin qui dure jusqu'à 2 minutes de plus pour en éviter un sur lequel il se ferait contrôler une fois avec l constant. En effet, sur un chemin de longueur L arcs avec un contrôle, nous avons $v_n = -4 \sum_{i=0}^I tt(k_i) - 8 - 0,2L$ et lorsque sur un chemin de longueur L , il n'y a pas de contrôle, nous avons $v'_n = -4 \sum_{i=0}^{i=I} tt(k'_i) - 0,2L$. Tant que $v'_n > v_n$, ce qui équivaut à $\sum_{i=0}^{i=I} tt(k'_i) < \sum_{i=0}^{i=I} tt(k_i) + 2$, le décideur choisira le chemin sans contrôle.

3.2.3 Spécification du terme d'erreur

En reprenant ce qui a été expliqué dans la section 2.1.2 de la revue de littérature, nous savons que le terme d'erreur est par définition la différence entre l'utilité que le décideur obtient et l'utilité observée par l'analyste et sera traité comme une variable aléatoire qui suit la loi de Gumbel et chaque terme d'erreur sera identiquement, indépendamment distribué (iid). D'après l'équation (2.30), la probabilité qu'un décideur n choisisse l'arc a , $a \in A(k)$ sachant qu'il se trouve en k est:

$$Pr_n^d(a|k) = \frac{e^{\frac{1}{\mu}(\beta_{tt}tt(a)+\beta_{cr}cr(a)+\beta_l l(a)+V_n^d(a))}}{\sum_{a' \in A(k)} e^{\frac{1}{\mu}(\beta_{tt}tt(a')+\beta_{cr}cr(a')+\beta_l l(a')+V_n^d(a'))}} \quad (3.5)$$

Cette formulation logit sera par la suite très pratique pour calculer la fonction de

log-vraisemblance $LL(\beta)$. L'espérance du maximum d'utilité de a vers la destination $V_n^d(a)$ est déterminée en résolvant l'équation de Bellman (2.29). Avant de terminer cette spécification de la fonction d'utilité déterministe, il est important de préciser que des attributs pertinents comme la proximité d'un point stratégique (aéroport, zone industrielle etc...), l'accès autorisé ou non d'une portion de route ou l'heure du parcours, ne figurent pas dans ce modèle mais les données pourraient être complétées pour de futures analyses. Nous allons maintenant pouvoir simuler des observations en fixant des valeurs à β_{tt} , β_{cr} et β_l . Dans le chapitre qui suit, nous allons montrer que notre modèle est valide puisque nous retrouvons les valeurs fixées des paramètres avec des observations complètes et partielles.

Chapitre 4

Résultats expérimentaux

Dans ce chapitre nous commencerons par présenter une méthodologie pour la validation de notre modèle grâce à des observations synthétiques. Bien que nous disposons d'une fonction avec des paramètres fixés et admissibles (voir la discussion dans la fin de la section 2.6.2), nous devons choisir une partie du réseau routier dans laquelle il sera facile de déterminer une paire origine-destination sachant le lieu d'interception du décideur par un agent de CRQ. Ceci nous amène à introduire la notion de sous-réseau. Avec ces éléments en main, nous pouvons maintenant simuler des observations qui sont des suites d'arcs ayant toutes la même origine et la même destination. Une fonction de log-vraisemblance associée à ces observations permettra de faire le travail d'estimation. Nous continuerons en montrant comment obtenir des observations partielles et comment nous avons obtenu une nouvelle fonction de log-vraisemblance en utilisant malgré tout une génération de routes. Nous terminerons en montrant que les valeurs numériques de nos estimations de paramètres ne sont pas significativement différentes des valeurs des paramètres du modèle de référence que ce soit dans le cas où les observations sont complètes ou dans l'autre dans une certaine limite.

4.1 Validation avec des observations synthétiques

Procédons à la validation de notre approche. L'objectif est de montrer que le modèle proposé peut être estimé avec un bon niveau de confiance sachant que les observations peuvent être partielles. Dans le chapitre précédent, nous avons présenté des attributs de réseau collectés sur le terrain, avec lesquels nous obtenons un graphe

représentant le vrai réseau du Québec. Notons que sur les observations du terrain, nous ne savons pas avec exactitude à quel endroit du réseau l'éventuelle infraction a été repérée puisque c'est le lieu où s'effectue le contrôle qui est enregistré et l'origine ainsi que la destination ne sont pas connues. Les paramètres estimés ne peuvent être exploitables que s'ils peuvent être comparés à des paramètres d'une fonction qui est non seulement du même type mais qui est aussi parfaitement déterminée. Pour étudier le comportement des camionneurs avec un modèle de choix discret, nous travaillerons par conséquent avec des observations synthétiques. D'abord, nous fixons une fonction de référence, dont les paramètres sont connus, avec laquelle nous simulerons des observations complètes. Estimer notre modèle avec des observations synthétiques répond doublement à nos attentes par le fait que nous pourrions à la fois faire des estimations en connaissant les valeurs des paramètres et mesurer les éventuels biais, ce qui n'est jamais possible avec les données réelles car le vrai modèle est inconnu. Nous utiliserons dans un premier temps des observations synthétiques constituées de routes complètes avec l'origine et la destination connues, puis nous continuerons avec des observations synthétiques partielles. Avec le premier échantillon d'observations, nous vérifions que les valeurs des paramètres estimés ne sont pas statistiquement différentes des valeurs des paramètres fixés par un modèle de référence. Avec le deuxième échantillon d'observations partielles dérivant du premier et dont des arcs sont progressivement retirés, nous allons vérifier que nous sommes bien capables d'identifier les paramètres et de montrer que leurs nouvelles valeurs ne sont pas significativement différentes des valeurs du modèle de référence, dans une certaine limite. Dans ce dernier cas, nous ne pouvons plus calculer la probabilité qu'une route soit choisie par le décideur. Nous pouvons en revanche nous inspirer des travaux faits dans (Bierlaire et Frejinger, 2008), expliqués dans la section 2.5.4 de la revue de littérature, et calculer cette probabilité en prenant la moyenne des probabilités des routes complètes qui partagent les mêmes arcs que la route incomplète considérée. Donc, pour faire le travail d'estimation, nous nous serons amenés à générer des routes complètes à partir de routes partielles. L'utilisation du modèle Logit Récursif nécessite la connaissance d'une paire origine-destination. Pour cette raison nous introduisons la notion de sous-réseau dans la section 4.1.1 pour définir avec certitude cette paire lorsque nous travaillons avec des routes incomplètes.

4.1.1 Le sous-réseau

Le réseau que nous modélisons correspond au réseau routier de la province du Québec. Le graphe orienté $G(\nu, A)$ est constitué d'un nombre d'arcs $|A| = 10320$ et de $|\nu| = 4946$ nœuds. Le réseau de Beloeil compte $|A| = 415$ arcs et $|\nu| = 224$ nœuds. Le choix de porter l'étude sur Beloeil s'explique par le fait que les calculs seront plus rapides en raison de la faible taille de la matrice d'incidence associée et nous pouvons étendre l'étude du modèle de choix de routes sur l'ensemble du Québec ensuite, puisque le modèle logit récursif n'est pas dépendant du réseau. Dans cette même logique, nous nous concentrerons sur une portion du réseau comportant un contrôle, des routes de contournement et sur laquelle nous savons quels arcs sont empruntés par les décideurs. Sur cette portion de réseau, nous devons être capables de déterminer avec certitude une paire origine-destination au sens près pour pouvoir travailler avec le modèle Logit Récursif. Cette portion de réseau, appelée sous-réseau joue ainsi un rôle de premier plan dans notre étude. Les figures 4.1 et 4.2 sont des exemples de sous-réseaux possédant les caractéristiques précédemment mentionnées.

Nous voulons calculer la probabilité qu'un décideur choisisse sa route, celle-ci passant par l'observation σ'_n qui est ici l'arc contrôlé a'_n symbolisé par un triangle jaune sur les figures 4.1 et 4.2. En observant le graphe, et en se concentrant sur les sommets situés dans un voisinage de a'_n , on peut déterminer avec certitude un sommet origine ν_{n1} symbolisé par un disque vert et un sommet destination ν_{n2} , symbolisé par un hexagone rouge sur les figures 4.1 et 4.2, ceci, afin de pouvoir tester notre approche. Une fois qu'on a ces sommets, on peut les associer chacun à un arc origine k_0 et un arc destination $d = k_I$ puisque le modèle se base sur des arcs. C'est grâce aux renseignements fournis par les agents de terrain de CRQ qu'il a été possible de déterminer des sous-réseaux comme indiqué sur les figures 4.1 et 4.2. La présence d'un poste de contrôle de type infrastructure sur l'autoroute A20 va pousser des décideurs à passer par un chemin de contournement, par exemple, en quittant l'autoroute pour rejoindre successivement la rue Bernard Pilon puis le chemin Trudeau, alors que ce dernier est interdit à certains types de véhicules et enfin ils rejoignent l'autoroute A20. Ceci est un comportement observé sur le terrain, mais nous ne savons rien sur les comportements non observés et il se peut que certains décideurs fassent des détours encore plus longs parce qu'ils savent que le chemin Trudeau est le théâtre de

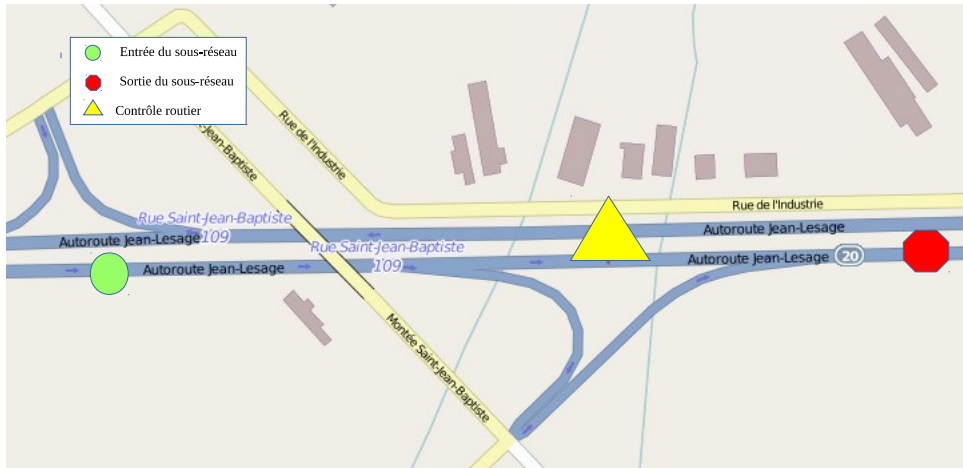


Figure 4.1 – Sous-réseau sur l'A20

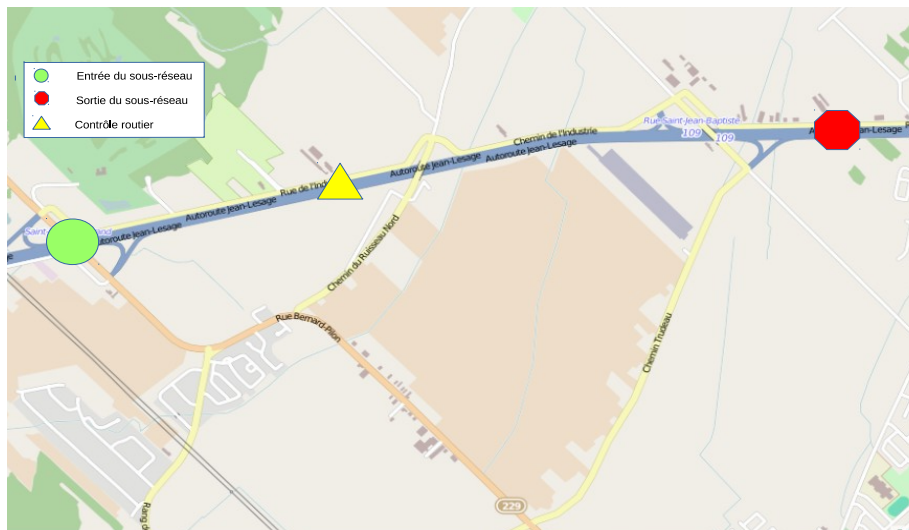


Figure 4.2 – Sous-réseau sur l'A20

contrôles fréquents. Mais lorsque nous travaillons avec des observations simulées, toutes les routes des décideurs passent par la même origine et la même destination. Le couple origine-destination est fixé à l'avance.

Définition:

Le sous-réseau $S_n(\sigma'_n, \nu_{n1}, \nu_{n2})$ est l'ensemble des routes dont l'origine est ν_{n1} , la destination est ν_{n2} et qui passent par σ'_n .

Notons que le sous-réseau grâce auquel nous connaissons avec certitude une paire origine-destination permet de tester notre modèle. Les portions de réseau étudiées illustrées par les figures 4.1 et 4.2 sont raisonnablement limitées pour éviter d'avoir

un trop grand nombre de routes possibles entre l'origine ν_{n1} et la destination ν_{n2} .

4.1.2 Génération des observations synthétiques complètes

Dans cette partie, nous générons N routes dont les origines et destinations seront respectivement ν_{n1} et ν_{n2} pour chaque décideur n . Nous utilisons un sous-réseau ayant seulement une origine et une destination. Plus tard, on pourra générer des routes à partir de plusieurs couples origine-destination, mais pour l'instant, on peut supprimer l'indice lié au décideur n pour alléger la notation de ces sommets. Les observations créées s'obtiennent en appliquant les résultats liés au modèle récursif logit. La fonction d'utilité est linéaire en β , dont les composantes seront fixées. Pour certaines valeurs de β , l'équation de Bellman (2.27) n'aura pas de solutions mais au bout de quelques tests, il est possible de trouver des valeurs admissibles. Une fois qu'un vecteur $\tilde{\beta}$ sera validé, il suffira d'exécuter un programme sous Matlab (conçu par Tien Anh Mai) pour avoir les N observations. Ensuite, il reste à calculer la fonction de log-vraisemblance et avec un module du même programme, il sera possible de faire nos estimations. Les valeurs numériques pour $\tilde{\beta}$ sont: $\tilde{\beta}_{tt} = -3,30$, $\tilde{\beta}_{cr} = -1,75$ et $\tilde{\beta}_l = -0,70$. La fonction d'utilité déterministe pour le modèle de référence s'écrit:

$$v_n(a) = -3,30tt(a) - 1,75cr(a) - 0,70l(a), \forall a \in A(k) \quad (4.1)$$

4.1.3 Expression de la log-vraisemblance

L'expression de la fonction de log-vraisemblance, notion abordée dans la revue de littérature, dans la section 2.4.1, pour un échantillon de N observations s'obtient directement à partir de (2.34).

$$LL(\beta) = \ln \prod_{n=1}^N P(\sigma_n), \quad (4.2)$$

$$= \frac{1}{\mu} \sum_{n=1}^N \sum_{i=0}^{I_n-1} (v(k_{i+1}|k_i) - V(k_0)) \quad (4.3)$$

Puisque nous avons suggéré un modèle où la fonction d'utilité déterministe est linéaire par rapport aux paramètres, nous pouvons expliciter la condition d'ordre 1 (2.17):

$$\frac{\partial LL(\beta)}{\partial \beta_q}(\widehat{\beta}_q) = \frac{1}{\mu} \sum_{n=1}^N \left[\sum_{i=0}^{I_n-1} x_{i+1|i}^q - \frac{\partial V(k_0^n)}{\partial \beta_q} \right], \quad q = tt, cr, l \quad (4.4)$$

$$= \frac{1}{\mu} \sum_{n=1}^N \left[x_{nq} - \frac{\partial V(k_0^n)}{\partial \beta_q} \right], \quad q = tt, cr, l \quad (4.5)$$

On remarquera que l'expression de la log-vraisemblance est facile à manipuler puisqu'elle fait intervenir directement la valeur des attributs. Le calcul de $\frac{\partial V(k_0^n)}{\partial \beta_q}$ s'obtient en différentiant (2.37). Ainsi, nous avons:

$$\frac{\partial \varphi}{\partial \beta_q} = (I - M)^{-1} \left[\frac{\partial M}{\partial \beta_q} \right] \varphi \quad (4.6)$$

Sachant que $\varphi_k = e^{\frac{1}{\mu} V(k)}$, $\frac{\partial \varphi}{\partial \beta_q} = \varphi \circ \frac{1}{\mu} \frac{\partial V}{\partial \beta}$ nous avons:

$$\frac{\partial V(k_0^n)}{\partial \beta_q} = \mu \frac{1}{\varphi_0} \left(\frac{\partial \varphi}{\partial \beta_q} \right)_0 \quad (4.7)$$

Où φ_0 et $\left(\frac{\partial \varphi}{\partial \beta_q} \right)_0$ qui sont des éléments correspondants à l'origine.

4.1.4 Validation de l'estimateur du maximum de vraisemblance

Dans la partie précédente, nous présenté l'estimateur du maximum de vraisemblance, dont les caractéristiques sont montrées dans la section 2.4.1 de la revue de littérature, en exploitant les spécificités de notre modèle. A présent, nous allons aborder la phase de validation, pendant laquelle nous vérifions si les valeurs des paramètres estimées $\hat{\beta}$ ne sont pas significativement différentes de celles des paramètres du modèle étalon qui nous sont connues $\tilde{\beta}$. Comme expliqué dans la section 2.4.3 nous faisons des tests de Student à un seuil de signification de 0,05 sur un échantillon d'observations de taille $N = 500$.

Nous souhaitons faire un test de conformité en testant $(H_0) : \hat{\beta}$ n'est pas significativement différent de $\tilde{\beta}$. On a alors $H_1: \hat{\beta}$ est significativement différent de $\tilde{\beta}$. On va tester (H_0) avec un test de Student pour un seuil de signification de 0,05 sur un échantillon de taille $N = 500$.

Les résultats que nous obtenons sont résumés dans le tableau IV dont chaque ligne donne les valeurs des estimations pour deux échantillons. Nous avons pour chaque

<i>attribut x</i>	$\widehat{\beta}_x$	$\widetilde{\beta}_x$	$\widehat{\sigma}_x$	<i>statistique t</i>	$\widehat{\beta}_x$	$\widetilde{\beta}_x$	$\widehat{\sigma}_x$	<i>statistique t</i>
<i>tt</i>	-3.45	-3,30	0.29	0.51	-3.28	-3,30	0.31	0.08
<i>cr</i>	-1.80	-1,75	0.10	0.53	-1.83	-1,75	0.10	0.76
<i>l</i>	-0.76	-0,70	0.05	1.15	-0.72	-0,70	0.05	0.35
<i>tt</i>	-3.17	-3,30	0.29	0.43	-3.08	-3,30	0.27	0.79
<i>cr</i>	-1.85	-1,75	0.10	0.98	-1.82	-1,75	0.10	0.72
<i>l</i>	-0.74	-0,70	0.05	0.87	-0.76	-0,70	0.05	1.20
<i>tt</i>	-3.43	-3,30	0.27	0.50	-3.06	-3,30	0.28	0.85
<i>cr</i>	-1.63	-1,75	0.09	1.34	-1.83	-1,75	0.10	0.80
<i>l</i>	-0.65	-0,70	0.04	1.17	-0.73	-0,70	0.05	0.54
<i>tt</i>	-3.67	-3,30	0.36	1.02	-3.41	-3,30	0.27	0.41
<i>cr</i>	-1.79	-1,75	0.10	0.40	-1.79	-1,75	0.09	0.50
<i>l</i>	-0.71	-0,70	0.05	0.19	-0.73	-0,70	0.03	1.29
<i>tt</i>	-3.37	-3,30	0.30	0.24	-3.37	-3,30	0.24	0.27
<i>cr</i>	-1.79	-1,75	0.10	0.41	-1.75	-1,75	0.10	0.01
<i>l</i>	-0.72	-0,70	0.05	0.38	-0.71	-0,70	0.05	0.18

Table IV – Validation de l'estimation des paramètres avec des observations complètes

attribut la valeur estimée du paramètre suivie de sa valeur de référence, de l'écart-type de l'échantillon ainsi que la statistique de Student. Nous voyons que pour chaque échantillon, la statistique de Student est inférieure en valeur absolue au 0,975 quantile de la loi $N(0, 1)$. On admet que d'après nos résultats, $\widehat{\beta}$ n'est pas significativement différent de la vraie valeur du vecteur de paramètres $\widetilde{\beta}$ du modèle étalon.

En utilisant des observations synthétiques complètes, nous avons été capables de montrer qu'il est possible d'estimer avec un seuil de signification de 0,05 notre modèle de choix de routes. La prochaine étape est de refaire ce travail de validation mais avec un jeu d'observations partielles et nous allons expliquer dans la prochaine section comment les générer. Ce travail est nécessaire car au lieu d'avoir des observations complètes, celles-ci seront tronquées et nous devons vérifier que les valeurs des paramètres estimés convergent vers des valeurs significativement proches. Ce sera le cas comme le montrent les résultats dans la section qui suit. Il est important de préciser que nous nous sommes limités à une réalisation de l'échantillon et que notre résultat est une illustration du fait que l'estimateur est consistant, ce dernier point étant démontré dans l'étude (Fosgerau, Frejinger et Karlström, 2013).

4.2 Observations synthétiques partielles

Les observations partielles sont des chemins incomplets sur lesquels nous ne connaissons que l'origine, la destination et quelques arcs empruntés. Le but est de se rapprocher des conditions réelles dans lesquelles l'analyste ne dispose que d'une seule information qui est le lieu de contrôle du véhicule du décideur par les agents de CRQ, c'est-à-dire un seul arc. L'origine et la destination sont des informations que nous déduisons d'après la topologie du réseau. En prenant une zone du réseau proche de l'arc observé, il est possible de déterminer avec certitude l'origine et la destination du décideur. Cette section est l'occasion de montrer comment nous générons ces routes partielles et aussi comment nous arrivons à faire l'estimation malgré tout, sachant toutefois qu'il y aura un prix à payer. En effet, il faudra générer des routes complètes pour pouvoir définir une fonction de log-vraisemblance avec des valeurs finies.

4.2.1 Génération des observations synthétiques partielles

Nous nous baserons sur des observations synthétiques générées complètes qui seraient obtenues dans des conditions idéales. Parmi ces routes, on peut facilement identifier celles qui sont considérées comme des routes de contournement puisqu'elles sont plus longues que les routes contrôlées et qu'elles contournent le contrôle comme dans l'exemple de la section 4.1.1. Une fois identifiées, nous leur enlèverons progressivement des arcs. Ces derniers doivent être des arcs sur lesquels l'attribut représentant la présence de contrôle a une valeur de 0. Ceux dont cette valeur est 1 sont conservés puisque sur le terrain, ils sont relevés lors des MTV et seront donc observés. On fait ce travail sur chaque observation complète et on obtient finalement notre ensemble d'observations incomplètes. Les routes qui ne sont pas considérées comme des routes de contournement sont considérées connues.

Travailler avec un échantillon de chemins incomplets a pour conséquence le fait qu'il n'est plus possible de calculer la fonction de log-vraisemblance. Il est alors nécessaire de redéfinir la probabilité qu'un décideur n choisisse une route qui n'est que partiellement observée σ'_n .

Ce que nous allons faire, c'est de faire en sorte de se retrouver dans une situation où on sait calculer la log-vraisemblance. Pour ça, on va devoir retravailler avec des chemins complets. Puisque nous avons au moins le couple origine-destination et un

arc, nous pouvons tenter une approche inspirée de (Bierlaire et Frejinger, 2008) et calculer une nouvelle fonction de log-vraisemblance, puis poursuivre en estimant de nouveau les paramètres de notre modèle de référence.

4.2.2 Fonction de log-vraisemblance modifiée

Considérons chaque observation partielle σ'_n . Nous savons que la vraisemblance d'une observation σ'_n est la probabilité que le décideur n passe par σ'_n , $\Pr(\sigma'_n)$. Si on reprend l'équation (2.25) en l'adaptant à notre situation, nous disons que la probabilité $\Pr(\sigma'_n)$ est la somme pondérée des probabilités sur l'ensemble des choix de chemins complets qui partagent les mêmes arcs que σ'_n en ayant la même origine ν_1 et la même destination ν_2 . Nous supposons que l'observation σ'_n a autant de chance d'être reconstituée par chacun des chemins σ_{nl} appartenant à l'ensemble des choix de chemins S_n et par conséquent, la pondération devra refléter le côté équiprobable. L'adaptation de (2.25) donne:

$$\Pr(\sigma'_n) = \sum_{\sigma_{nl} \in S_n} \Pr_n(\sigma'_n | \sigma_{nl}) \Pr_n(\sigma_{nl} | S_n) \quad (4.8)$$

$\Pr_n(\sigma_{nl} | S_n)$ étant calculé en utilisant les résultats du modèle de choix de routes logit récursif. Par souci de simplicité nous notons $S_n(\sigma'_n, s) = S_n$, (ν_1, ν_2) formant le couple origine-destination s . Étant donné l'hypothèse d'équiprobabilité, nous avons:

$$\Pr(\sigma'_n) = \frac{1}{|S_n|} \sum_{\sigma_{nl} \in S_n} \Pr_n(\sigma_{nl} | S_n) \quad (4.9)$$

En considérant les N observations, la fonction de vraisemblance est:

$$L(\beta) = \prod_{n=1}^N \frac{1}{|S_n|} \sum_{\sigma_{nl} \in S_n} \Pr_n(\sigma_{nl} | S_n) \quad (4.10)$$

Ce qui donne une fonction de log-vraisemblance:

$$LL(\beta) = \sum_{n=1}^N \ln \left(\frac{1}{|S_n|} \sum_{\sigma_{nl} \in S_n} \Pr_n(\sigma_{nl} | S_n) \right) \quad (4.11)$$

La condition de premier ordre est:

$$\frac{\partial LL(\beta)}{\partial \beta_q}(\widehat{\beta}_q) = \frac{\partial}{\partial \beta_q} \sum_{n=1}^N \ln \sum_{\sigma_{nl} \in S_n} \left(e^{-\frac{1}{\mu} V(k_0)} \prod_{i=0}^{I-1} e^{\frac{1}{\mu} (v_{nl}(k_{i+1}|k_i))} \right) \quad (4.12)$$

$$= \frac{\partial}{\partial \beta_q} \sum_{n=1}^N \ln \left(e^{-\frac{1}{\mu} V(k_0)} \sum_{\sigma_{nl} \in S_n} e^{\frac{1}{\mu} \sum_{i=0}^{I-1} (v_{nl}(k_{i+1}|k_i))} \right) \quad (4.13)$$

$$= \sum_{n=1}^N \frac{\partial}{\partial \beta_q} \ln \left(e^{-\frac{1}{\mu} V(k_0)} \sum_{\sigma_{nl} \in S_n} e^{\frac{1}{\mu} v(\sigma_{nl})} \right) \quad (4.14)$$

$$= \sum_{n=1}^N \left[-\frac{1}{\mu} \frac{\partial V(k_0)}{\partial \beta_q} + \frac{\partial}{\partial \beta_q} \ln \sum_{\sigma_{nl} \in S_n} e^{\frac{1}{\mu} v(\sigma_{nl})} \right] \quad (4.15)$$

$$= \frac{1}{\mu} \sum_{n=1}^N \left[\frac{\sum_{\sigma_{nl} \in S_n} x_{nlq} e^{\hat{\beta}^T x_{nl}}}{\sum_{\sigma_{nl} \in S_n} e^{\hat{\beta}^T x_{nl}}} - \frac{\partial V(k_0)}{\partial \beta_q} \right] = 0 \quad (4.16)$$

Lorsque $\sigma'_n = \sigma_n$, nous avons $S_n = \{\sigma_n\}$, $|S_n| = 1$ et on retrouve la relation de premier ordre (4.4) pour le cas particulier où l'analyste dispose d'une observation complète.

Un des atouts du modèle logit récursif est le fait qu'il n'impose pas de travailler avec des restrictions sur l'ensemble des alternatives alors que maintenant, nous allons devoir composer des routes lors de la définition d'un ensemble S_n pour chaque observation partielle. Pour des travaux de recherche à venir, une autre formulation ne nécessitant pas de générer des chemins devrait être faite. Le fait de travailler dans un premier temps en générant un ensemble de routes a l'avantage de réduire la complexité et de nous permettre de faire une première investigation du problème d'identification des paramètres. Si nous n'y arrivons pas dans ce cas, nous n'y arriverons pas dans l'autre.

4.2.3 Validation de l'estimateur du maximum de vraisemblance

Les observations synthétiques incomplètes σ'_n s'obtiennent en enlevant des arcs aux observations synthétiques complètes σ_n . Considérons la figure 4.3. Sur la configuration (a) nous avons deux observations complètes avec la route contrôlée qui est au dessus alors que la route de contournement est en dessous. Nous enlevons à cette dernière une suite d'arcs, à partir de a_n et elle devient une observation partielle. Sur la configuration (b), nous sélectionnons 2 suites d'arcs, qui vont compléter l'observation partielle pour former 2 observations complètes, $|S_n| = 2$. La configura-

tion (c) indique que nous continuons d'enlever des arcs jusqu'à la destination et nous sélectionnons 10 suites d'arcs pour former 10 observations complètes, $|S_n| = 10$. La dernière configuration ne nous permet d'observer qu'un seul arc en plus de l'origine et de la destination. Ce dernier cas de figure est celui qui nous rapproche le plus des observations réelles et le nombre de routes passant par a_n est $|S_n| = 20$.

La sélection des suites d'arcs servant à compléter les observations partielles se fait en se basant sur les observations complètes. Les échantillons d'observation que nous faisons ne sont pas aléatoires et nous devons introduire un correctif pour le biais résultant.

Pour les routes de contournement, $|S_n|$ a une valeur dépendant des arcs enlevés et est donc dépendante du réseau. Faire l'analyse du choix de routes sur une autre partie du réseau entraînera des valeurs différentes pour $|S_n|$. Le tableau V montre les résultats des estimations en fonction du nombre d'alternatives dans le sous-réseau. Les trois premières lignes montrent que, pour un échantillon témoin complet correspondant à la configuration (a) de la figure 4.3, la statistique t vaut 1, 1, 0, 70 et 0, 31 pour respectivement le paramètre associé à tt , cr et l . En ce qui concerne la configuration (b), la statistique t vaut cette fois 0, 35, 0, 59 et 0, 26. Finalement les trois lignes restantes correspondant à la configuration (c) nous donnent les valeurs de la statistique t qui sont 0, 22, 0, 32 et 1, 39. Ces valeurs sont toutes inférieures à 1, 96. Dans les cas illustrés par les configurations (b) et (c) les valeurs des estimés ne sont donc pas statistiquement différentes des valeurs réelles des paramètres. Si nous enlevons trop d'arcs pour ne garder qu'un arc observé a_n comme dans la configuration (d), nous n'avons plus de convergence pour les estimés. En effet, les valeurs de la fonction de log-vraisemblance varient infinitésimalement entre deux itérations alors que la valeur absolue du gradient restera supérieure à zéro. L'algorithme se termine, au bout du nombre maximum d'itérations que nous fixons, avec un échec dans l'estimation. Cette configuration est telle que les observations ne contiennent plus suffisamment d'information pour identifier les paramètres même lorsque le modèle est bien spécifié, ce qui n'est pas le cas avec des données réelles. Nous avons réussi à estimer notre modèle sachant que les observations sont incomplètes, mais avec une limite dans le manque d'information. Il faut continuer les tests avec d'autres routes à intégrer dans le sous-réseau car les résultats finaux dépendent de l'échantillonnage des routes σ_{nl} .

Les résultats dans cette expérience montrent qu'on peut identifier des paramètres

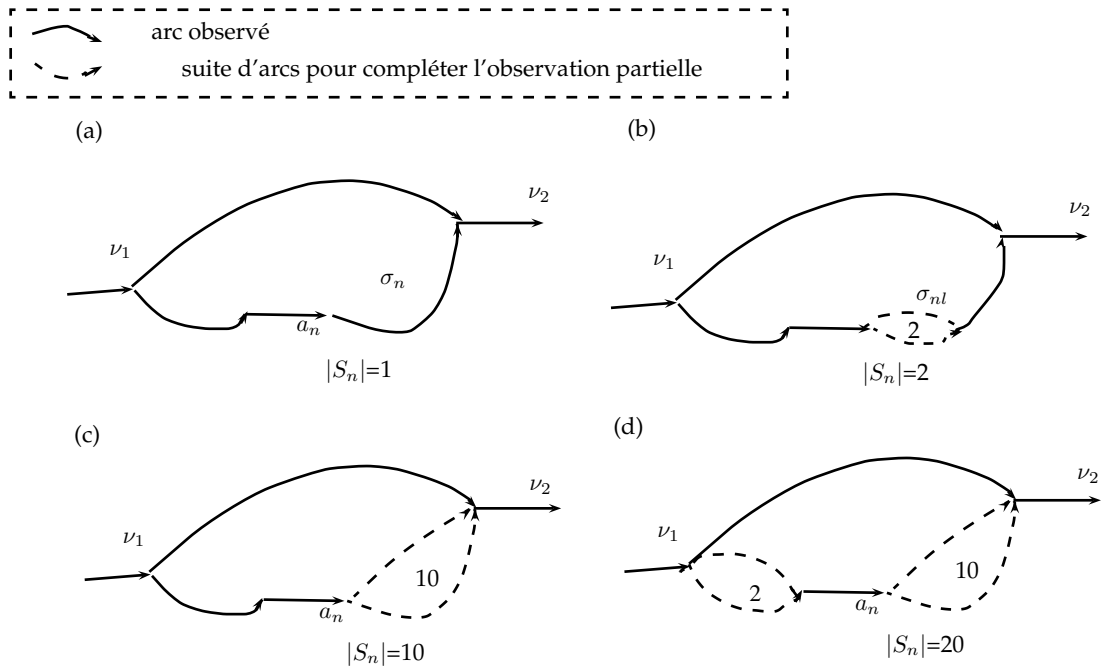


Figure 4.3 – Observations partielles issues d'observations complètes

alors que les observations sont incomplètes. Le moment où nous devons travailler avec des observations réelles, il faudra qu'elles contiennent plus d'information qu'elles n'en ont actuellement. Comme pour toute démarche expérimentale, il y a certainement des points à améliorer et il faut essayer de prendre du recul sur nos résultats. Notre expérience est limitée dans le fait que nous utilisons un échantillon et nous ne pouvons qu'affiner nos résultats en augmentant le nombre d'échantillons d'observations pour des investigations futures. Le chapitre suivant nous propose une réflexion sur la pertinence de notre étude des comportements des usagers sur un réseau routier et sur des pistes à prendre pour de prochaines études à ce sujet.

<i>attribut x</i>	$\widehat{\beta}_x$	$\widetilde{\beta}_x$	$\widehat{\sigma}_x$	<i>statistique t </i>	<i>Remarques</i>
<i>tt</i>	-3.02	-3,3	0.25	1.11	$ S_n = 1$
<i>cr</i>	-1.69	-1,75	0.09	0.70	
<i>l</i>	-0.71	-0,7	0.05	0.31	
<i>tt</i>	-3.41	-3,3	0.31	0.35	$ S_n = 2$
<i>cr</i>	-1.69	-1,75	0.10	0.59	
<i>l</i>	-0.71	-0,7	0.05	0.26	
<i>tt</i>	-3.24	-3,3	0.27	0.22	$ S_n = 10$
<i>cr</i>	-1.78	-1,75	0.09	0.32	
<i>l</i>	-0.77	-0,7	0.05	1.39	
<i>tt, cr, l</i>	<i>pas de convergence</i>				$ S_n = 20$

Table V – Validation de l'estimation des paramètres avec des observations partielles

Chapitre 5

Discussion

L'étude du comportement des usagers sur les réseaux routiers passe par l'analyse de leurs choix de routes. Parmi la variété de modèles qui permettent de faire de telles analyses, nous avons porté notre attention sur les modèles de choix discrets. S'il fallait justifier cette décision, nous dirions que d'un côté, le choix de routes est par nature un choix discret et de l'autre nous avons comme objectif futur d'implémenter un outil d'aide à la décision capable de prédire des flux dans les réseaux basé sur ce type de modèle.

Dans la littérature, nous ne trouverons probablement pas de sources proposant une hiérarchie des modèles et il n'existe pas de modèles parfaits pour une situation donnée. Notre démarche est guidée par la recherche d'un bon compromis entre la complexité des calculs et une simulation la plus proche possible des phénomènes observés ce qui nous a amenés à choisir le modèle logit récursif de ([Fosgerau, Frejinger et Karlström, 2013](#)). Deux approches se distinguent dans la modélisation des choix discrets par le fait que la première propose un choix de routes statique contrairement à la deuxième dans laquelle le choix se fait de manière dynamique. Le logit récursif que nous avons présenté dans le chapitre 2 se situe dans cette dernière catégorie dont les avantages sont les suivants: il n'est plus nécessaire d'élaborer une stratégie d'échantillonnage de routes pour établir les ensembles d'alternatives, ce qui nous épargne des efforts supplémentaires de calculs et des biais dans l'estimation. De plus, la possibilité de prédire des flux aura son importance pour de futures études. Le problème posé par les corrélations de routes peut se traiter par un correctif mais cette solution ne s'applique pas sans efforts supplémentaires en calculs. Les paramètres sont relativement faciles à estimer ceci étant dû à l'expression simple de la fonc-

tion de log-vraisemblance. Cette estimation, issue de la résolution d'un problème d'optimisation non linéaire sans contraintes donne un EMV sans biais, convergent en probabilité et efficace.

Le principal défi dans ce projet a été de travailler avec des observations partielles. Nous avons abordé l'étude du comportement des camionneurs en sachant donc quel modèle on voulait utiliser et nous avons commencé par réfléchir avec quelles données et quelles observations il faudrait travailler. Le travail de description des données permettra à d'autres de poursuivre ce que nous avons fait, leur donnera une vue claire sur ce qui est disponible. Sûrement, il y aura des idées pour collecter de nouvelles données. Ce travail a nécessité une capacité de communication car il a fallu identifier des personnes qui détenaient potentiellement des informations, puis les obtenir. Il a aussi permis de mettre en relief la nécessaire collaboration entre les chercheurs et sans surprises, on constate que travailler en équipe ou autour d'une équipe est toujours bénéfique.

Avec les attributs collectés, nous avons pu modéliser le réseau routier, puis nous avons proposé une manière de calculer la fonction qui permettra de prédire les probabilités de choix de routes d'un décideur. Le modèle étant spécifié, nous pouvions commencer à étudier les comportements grâce à des comportements observés ou plus simplement, des observations. Concernant ces dernières, nous savions qu'il serait difficile d'en avoir qui soient exploitables, mais ce n'était pas un obstacle insurmontable dans notre objectif d'étude du comportement des camionneurs. L'idée d'utiliser des observations synthétiques est venue rapidement, car elle présentait les avantages de faire une validation de notre modèle et de mesurer avec précision un biais puisque nous avons la maîtrise de la fonction décrivant le comportement. Ces simulations constituent une étape essentielle dans la mesure où elles ont montré qu'il est possible de relever le défi de faire des estimations avec des observations partielles sur un vrai réseau. Bien que nous soyons capables de faire des inférences avec un seuil de signification de 0,05, ces estimations nécessitent, pour augmenter nos chances de succès, d'observer une quantité minimum d'arcs. Il est très probable que ce soit le cas avec les données collectées sur le terrain. Aussi, nous pouvons penser à faire une expérience Monté-Carlo pour investiguer les résultats sur plus d'échantillons.

Il est difficile de faire des extrapolations de nos résultats avec les données du terrain, car nous avons introduit nous-mêmes une origine et une destination et que nous avons imposé une fonction expliquant le comportement, celle-ci étant inconnue

en réalité. Cette approche, bien que nous éloignant des phénomènes observés dans la réalité, nous permet de tirer des conclusions concrètes. Ainsi, nous pouvons suggérer des améliorations dans la collecte de données avec au minimum la saisie de l'origine et de la destination ainsi que le lieu exact de l'interpellation pour chaque MTV. On pourrait suggérer aussi de relever les heures pour une étude des effets de saisonnalité. Une telle modification du questionnaire utilisé lors des contrôles permettrait au CRQ de mieux cerner le comportement des camionneurs et finalement, les prises de décision concernant le déploiement des ressources de contrôle seraient optimisées.

Dans notre problème de choix de routes, les décideurs sont au courant des emplacements des contrôles et les observations que nous aurions concerneraient seulement les décideurs qui ont choisi une portion de route contrôlée. L'échantillon des observations comporte alors un biais de sélection. Il existe des moyens mis en œuvre pour corriger ce biais, tel que présenté dans (Heckman, 1979) ou dans (Manski, 1977). Dans ce dernier article, nous relevons que lorsque l'analyste procède à un échantillonnage exogène, il tire un échantillon d'individus et ensuite observe leurs choix et c'est un échantillonnage endogène lorsqu'à partir d'un ensemble de choix, il va observer les caractéristiques des individus et des alternatives. Puis, si l'analyste dispose d'une constante spécifique à chaque alternative dans sa représentation de l'utilité, alors un modèle logit peut être estimé comme si l'échantillon était exogène alors qu'il est purement endogène, grâce à l'utilisation d'un logiciel d'estimation du maximum de vraisemblance basé sur l'échantillonnage exogène, sachant que ces constantes vont absorber les biais. Un correcteur post-estimation de biais peut être mis en place si on connaît les parts de marché, c'est à dire si on connaît les proportions de décideurs pour chaque alternative. Ce correctif est tel que:

$$E(\hat{\alpha}_j^*) = \alpha_j^* - \ln(A_j/S_j) \quad (5.1)$$

où $\hat{\alpha}_j^*$ est l'estimé de la constante pour l'alternative j , α_j^* est la valeur réelle de la constante, A_j est la part de décideurs dans la population à choisir j et S_j est la part des décideurs ayant choisi j dans l'échantillon de choix. Dans notre contexte, il faut connaître la proportion de camionneurs, dans toute la population de camionneurs, qui choisissent un arc contrôlé. Cette restriction sur la connaissance du marché sera levée dans (Bierlaire, Bolduc et McFadden, 2008). Cependant, le modèle de choix discret doit appartenir à une classe appelée extremum généralisé en blocs-additifs et

la fonction d'utilité déterministe doit inclure toutes les constantes requises.

De plus, comme autre difficulté, notons que chaque décideur associe une fonction d'utilité aux alternatives à sa manière et cette fonction nous est inaccessible. Par conséquent, notre fonction d'utilité déterministe qui décrit notre représentation de la réalité sera plus difficile à spécifier. Il se peut aussi que le logit ne soit pas le bon modèle. Une étude sur les classes latentes serait clairement à envisager pour mieux rendre compte de la réalité. Puis, que ce soit avec des données réelles ou synthétiques, si elles sont incomplètes, nous avons vu que déterminer un sous-réseau implique une génération de routes nécessaire pour calculer les valeurs des log-vraisemblances. Cette approche vient donc contrebalancer l'avantage principal du modèle logit récursif et vient ajouter un biais supplémentaire.

La théorie de l'utilité aléatoire nous a offert un cadre pour l'étude des choix de routes. Cependant, d'autres méthodes ont inspiré des chercheurs. Ainsi, on peut faire référence à l'étude de ([Dougherty, 1995](#)), qui offre une revue des méthodes se basant sur les réseaux neuronaux, ainsi qu'à l'étude de ([Yamamoto, Kitamura et Fujii, 2002](#)) qui utilise des arbres de décision pour modéliser des choix de routes quand deux routes sont disponibles tels que mentionnés dans la thèse suivante ([Frejinger, 2008](#)). Alors que nous essayons de résoudre des problèmes de choix de routes, il va sans dire que les modèles de choix discrets peuvent couvrir une large gamme de problèmes. Une revue des applications des modèles de choix discrets est disponible dans les travaux de McFadden, 1982, Manski, 1981 et d'Amemiya, 1981.

Les patrouilles de CRQ doivent alors s'adapter et changer de position en conséquence. Sachant que les décideurs et les gestionnaires du réseau ont des objectifs opposés, on pourrait concevoir une suite à ce projet dans lequel l'analyste étudierait les équilibres avec la théorie des jeux, particulièrement la théorie des programmes bi-niveaux.

Pouvoir tester des algorithmes pour générer des routes pour traiter les observations partielles est aussi envisageable mais idéalement, il serait préférable de rester cohérent avec le modèle logit récursif utilisé ici. D'autres méthodes pour traiter des observations partielles dans des processus de prise de décisions dans les chaînes de Markov sont exposées dans ([Lovejoy, 1991](#)).

Ce mémoire a été l'occasion d'investir des connaissances théoriques dans un problème rencontré dans le monde industriel. Travailler avec un réseau routier réel pose des difficultés supplémentaires par rapport au fait d'utiliser des réseaux simplifiés

faits par soi-même car il a fallu effectuer un travail de collecte de données liées au réseau sachant qu'on est jamais à l'abri d'erreurs. Le graphe modélisant le réseau n'est pas fortement connexe et il a été nécessaire de restreindre l'analyse sur les sous-graphes fortement connexes pour éviter des erreurs numériques et faire en sorte que notre approche reste valide. Bien que les observations utilisées soient synthétiques, c'est avec le réseau réel et des attributs réels que les estimations sont réalisées, ce qui rend les analyses et les résultats décrits dans ce mémoire d'autant plus pertinents.

Références bibliographiques

- Abdel-Aty, M.A., Kitamura, R., Jovanis, P.P, Reddy, P, Vaughin, K.M, 1995. New approach to route choice data collection: Multiphase, computer-aided telephone interview panel surveys using geographic information systems data base. *Transportation Research Record* 1493, pp 159–169.
- Albright, R., Manski, C.F, Lerman, S.R., 1977. Report on the Development of an Estimation Program for the Multinomial Probit Model. Prepared for Federal Highway Administration, Washington D.C.
- Amemiya, T., 1981. Qualitative Response Models: A Survey. *Journal of Economic Literature*, pp 1483–1536.
- Baillon, J.B., Cominetti, R., 2008. Markovian traffic equilibrium. *Mathematical Programming* 111(1-2), pp 33–56.
- Bekhor, S., Ben-Akiva, M., Ramming, M., 2002. Adaptation of logit kernel to route choice situation. *Transportation Research Record* 1805(1), pp 78–85.
- Bellman, R., 1957. *Dynamic Programming*. Princeton University Press, Princeton.
- Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short-term travel decisions. Hall, R. Ed., *Handbook of Transportation Science*, Kluwer, pp 5–34.
- Ben-Akiva, M., Lerman, S.R., 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Massachusetts.
- Bierlaire, M., Bolduc, D., McFadden, D., 2008. The estimation of Generalized Extreme Value models from choice-based samples. *Transportation Research Part B* 42(4), pp 381–394.

- Bierlaire, M., Frejinger, E., 2008. Route choice modeling with network-free data. *Transportation Research Part C* 16(2), pp 187–198.
- Bovy, P.H.L, Stern, E., 1990. *Route choice: wayfinding in transport networks*. Kluwer Academic Publishers.
- Burrell, J. 1968. Multipath route assignment and its application to capacity restraint. *Proceeding of the 4th International Symposium on the Theory of Road and Traffic Flow*.
- Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. Lesort, J.B. Ed., *Proceedings of the 13th international symposium on transportation and traffic theory*, Lyon, France.
- Daganzo, C., Bouthelier, F., Sheffi, Y., 1977. Multinomial Probit and Quantitative Choice: A Computationally Efficient Algorithm. *Trans. Sci.* 11(1), pp 338–358.
- Dial, R., 1971. A probabilistic multi path traffic assignment algorithm which obviates path enumeration. *Transportation Research Part B* 5(2), pp 83–111.
- Dougherty, M, 1995. A review of neural networks applied to transport. *Transportation Research Part C: Emerging Technologies* 3(4), pp 247–260.
- Dutt, J., 1976. Numerical Aspects of Multivariate Normal Probabilities in Econometric Models. *Ann. Econ. Social Measurement* 5(4), pp 547–561.
- Fosgerau, M., Frejinger, E., Karlström, A., 2013. A link based network route choice model with unrestricted choice set. *Transportation Research Part B* 56(1), pp 70–80
- Frejinger, E., 2008. *Route choice analysis: data, models, algorithms and applications*. PhD. Thesis, EPFL.
- Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B* 41(3), pp 363–378.
- Gendron, B., Crainic, T.G, Florian, M., Metnani, A., Deslauriers-Gaboury, O., 2013. *Optimisation de la couverture du territoire contrôle routier Québec rapport d'étape 6*.

- Hausman, J., Wise, D., 1978. A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica* 46(1), pp 403–426.
- Heckman, J.T, 1979. Selection Bias as a Specification Error. *Econometrica* 47(1), pp 153–161.
- Kmenta J., 1971. *Elements of Econometrics*. Mcmillan, New York.
- Lovejoy, W.S, 1991. A survey of algorithmic methods for partially observed Markov decision process. *Annals of Operations Research* 28(1), pp 47–66.
- Luce, R., 1959. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New-York.
- Luce, R., Suppes, P., 1965. Preference, Utility and Subjective Probability. *Handbook of Mathematical Psychology*. In R. Luce, R. Bush and E. Galanter, Eds. Wiley, New York 3(1), pp 252–410.
- Mahmassani, H.S., Joseph, T., Jou, R.C, 1993. Survey approach for study of urban commuter choice dynamics. *Transportation Research Record* 1412(1), pp 80–89.
- Mai-Anh, T., 2012. *Revisiting Optimization Algorithms for Maximum Likelihood Estimation*. M.Sc.Thesis. Université de Montréal.
- Manski, C.F., 1977. The Structure of Random Utility Models. *Theory and Decision* 8(1), pp 229–254.
- Manski, C.F., 1981. *Structural Models for Discrete Data*. *Sociological Methodology*, pp 58–109.
- Manski, C.F., Lerman, S.R., 1977. The estimation of choice probabilities from choice based samples. *Econometrica* 45(8), pp 1977–1988.
- McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*. P. Zarembka, Eds. Academic Press, New York, pp 105–142.
- McFadden, D., 1978. Modeling the choice of residential location, in Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. Eds., *Spatial Interaction Theory And Planning Models*, pp 75–96. Reprinted in J. Quigley Ed., *The Economics Of Housing*, London: Edward Elgar 1(1), pp 531–552.

- McFadden, D., 1982. Econometric Analysis of Qualitative Response Models., W. Hildenbrand (ed.), *Advances In Econometrics: Invited Papers For The Fourth World Congress Of The Econometric Society*, Econometric Society Monograph, Cambridge, pp 1–37.
- Prato, C.G, 2004. Latent Factors and Route Choice Behaviour. PhD thesis, Politecnico di Torino.
- Tarjan, R. E., 1972. Depth-first search and linear graph algorithms. *SIAM Journal on Computing* 1(2), pp 146–160.
- Thurston, L., 1927. A Law of Comparative Judgment. *Psychological Review* 34(4), pp 273–286.
- Train, K., 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Vovsha, P., Bekhor, S., 1998. Link-nested logit model of route choice overcoming route overlapping problem. *Transportation Research Record* 1645(1), pp 133–142.
- Yamamoto, T., Kitamura, R., Fujii, J., 2002. Drivers route choice behavior: Analysis by data mining algorithms. *Transportation Research Record* 1807(1), pp 59–66.