



Institut de Statistique, Biostatistique et Actuariat

Traitement de la dualité incertitude/risque présente dans les
données télématiques par une approche "fuzzy theory" :
application au designing d'un produit d'assurance RC auto.

Membres du jury :

Prof. Ars Pierre *Promoteur*
Prof. Van Der Spiegel Jan *Co-promoteur*
Prof. Denuit Michel *Lecteur*

Mémoire présenté en vue de
l'obtention du master
en sciences actuarielles
par :

Derclaye Fanny

Louvain-La-Neuve
Janvier 2023

Résumé

Dans ce papier, un produit télématique d'assurance automobile a été conçu sur base de scores flous de conduite. Ces scores flous dépendent de données télématiques relatives au comportement routier qui sont récoltées à distance afin de modifier la tarification des assurances. La nature floue qui caractérise ces données a été modélisée à l'aide de la théorie des ensembles flous. Cette théorie tient compte de l'imprécision relative à certaines classes d'objets et permet l'adhésion graduelle d'un objet à un ensemble. De nombreux concepts ainsi que certaines définitions liées aux statistiques floues ont été définis pour cette théorie. La logique floue, qui fait partie de la théorie des ensembles flous, permet de prendre une décision en tenant compte de l'imprécision et de l'incertitude. C'est grâce à celle-ci que plusieurs sous-scores flous de conduite compris entre 0 et 100 ont été construits, chaque sous-score étant lié à un domaine bien précis (vitesse, prise de virages ...). Au plus le sous-score d'un assuré se rapproche de 100, au moins son comportement est perçu comme risqué en ce qui concerne la sinistralité. Ces sous-scores ont ensuite été utilisés au sein d'un système bonus-malus pour construire un produit télématique.

La source première de ce travail est un ouvrage de [Wierman \(2010\)](#), intitulé *An introduction to the mathematics of uncertainty*.

Abstract

In this paper, a telematics-based car insurance product has been designed based on fuzzy driving scores. These fuzzy scores depend on telematics data related to driving behavior that are collected remotely in order to modify insurance pricing. The fuzzy nature of this data was modeled using fuzzy set theory. This theory takes into account the imprecision relative to certain classes of objects and allows the gradual adhesion of an object to a set. Many concepts and some definitions related to fuzzy statistics have been defined for this theory. Fuzzy logic, which is part of the theory of fuzzy sets, makes it possible to make a decision taking into account imprecision and uncertainty. It is thanks to this theory that several fuzzy driving sub-scores between 0 and 100 have been built, each sub-score being linked to a specific domain (speed, cornering ...). The closer an insured's sub-score is to 100, the less risky his or her behavior is perceived to be in regard to the claims experience. These sub-scores were then used within a bonus-malus system to build a telematics product.

The primary source for this work is a book by [Wierman \(2010\)](#), entitled *An introduction to the mathematics of uncertainty*.

Avant-propos

Ce mémoire a été soumis en vue de satisfaire partiellement aux exigences du master en sciences actuarielles délivré par l'Université Catholique de Louvain.

Remerciements

Je tiens avant tout à remercier Pr. Pierre Ars et Jan Van Der Spiegel pour tout le temps qu'ils m'ont consacré pour l'élaboration de ce mémoire. Leurs idées m'ont éclairée et grandement aidée tout au long de la rédaction de ce mémoire.

Je souhaite également remercier de tout cœur mes parents pour le soutien qu'ils m'ont tant donné tout au long de mon parcours universitaire.

Un grand merci à ma grand-mère, Nicole, qui a pris le temps de relire chaque ligne de ce mémoire pour m'aider dans la correction de la syntaxe et de la formulation.

Finalement, je souhaite remercier mon grand frère, Tom, qui m'a donné de nombreux conseils pour la rédaction de ce mémoire.

Table des matières

Introduction	1
1 La tarification de l'assurance automobile et l'apport des données télématiques	3
1.1 Méthode classique de tarification de type GLM	3
1.1.1 Théorie du modèle	3
1.1.2 Hypothèses de base du modèle	4
1.1.3 Le cas des données de comptage	4
1.1.4 Prise en compte de l'exposition	5
1.2 L'apport des données télématiques dans l'assurance automobile	5
1.2.1 Le questionnement de l'assurance automobile traditionnelle	5
1.2.2 L'utilisation des données télématiques	6
1.2.3 Comportements de conduite liés aux accidents	8
1.2.4 Produits télématiques existants sur le marché de l'assurance automobile belge	8
1.3 Intégration de la télématique au sein du GLM	8
1.3.1 Score de conduite construit grâce à la télématique	8
1.3.2 Changements de comportement et gamification de l'assurance	9
1.3.3 La caractéristique floue des données télématiques	9
2 Théorie des ensembles flous	11
2.1 Définition d'un ensemble flou	11
2.2 Opérations sur les ensembles flous	13
2.2.1 Sous-ensemble flou	14
2.2.2 Intersection	14
2.2.3 Union	14
2.2.4 Complément	14
2.3 Propriétés d'un ensemble flou	15
2.3.1 La cardinalité scalaire	15
2.3.2 L'alpha coupe (α -coupe)	15
2.3.3 Le support	16
2.3.4 Le noyau (ou pic, mode)	16
2.3.5 La hauteur	16

2.3.6	Le socle	16
2.4	Nombres flous	17
2.4.1	Nombres flous triangulaires	18
2.4.2	Nombres flous trapézoïdales	18
2.5	L'arithmétique floue	19
2.5.1	Le principe d'extension de Zadeh	19
2.5.2	Application du principe d'extension aux statistiques flous	20
2.5.3	Distance entre deux nombres flous	21
3	La logique floue	23
3.1	Les systèmes d'inférence floue	23
3.1.1	Création des ensembles flous et des fonctions d'appartenance	24
3.1.2	Fuzzification	25
3.1.3	Règles floues	26
3.1.3.1	Règles multiples	26
3.1.3.2	Antécédents multiples	26
3.1.3.3	Agrégation de la sortie des règles	27
3.1.4	Défuzzification par la méthode du centre de gravité (COG)	28
3.2	Performance de la théorie	29
3.2.1	Avantages	29
3.2.2	Inconvénients	30
3.3	L'utilisation de la logique floue en assurance	30
4	Statistiques floues et application sur une variable télématique	31
4.1	Classification floue selon Gonzalez et De Campos	31
4.1.1	Théorie	31
4.1.2	Application : choix de l'assuré concernant le type de contrat	33
4.1.2.1	Prime pure d'un contrat traditionnel	33
4.1.2.2	Prime pure d'un contrat télématique	34
4.2	L'espérance floue selon Dubois et Prade	37
4.2.1	Théorie	37
4.2.2	Lien avec la théorie présentée par Gonzalez et de Campos via la mesure de Lebesgue	38
4.2.3	Application : approche floue des intervalles de confiance	40
4.3	L'espérance floue selon Kandel	42
4.4	L'espérance floue dans un espace de crédibilité	43
4.5	L'espérance floue selon Puri et Ralescu	44
5	Construction d'un produit télématique	45
5.1	Base de données	45
5.2	Construction de plusieurs sous-scores flous de conduite à l'aide de la logique floue	45
5.2.1	Corrélation et analyse en composantes principales (ACP)	46

5.2.2	Lien entre les regroupements effectués par l'ACP et la sinistralité	47
5.2.3	Création d'un système d'inférence floue pour chacun des sous-scores déterminés	48
5.3	Conception du produit télématique : le système bonus-malus	49
5.3.1	Mise en place du système	49
5.3.2	Exemple sur base de deux sous-scores	51
5.3.3	Équilibre du portefeuille	52
5.4	Perspective future de l'utilisation de l'approche fuzzy	52
	Conclusion	52
	Annexes	59
	A Présentation de la base de données	61
	B Présentation du package FuzzyR	65
	C Complément du chapitre 4	67
	D Complément du chapitre 5	69

Table des figures

2.1.1 Schéma descriptif d'un ensemble flou	13
2.2.1 Schéma des opérations sur les ensembles flous	15
2.3.1 Propriétés d'un ensemble flou	17
2.4.1 Nombres flous d'une forme triangulaire (gauche) et d'une forme trapézoïdale (droite)	18
3.1.1 Sous-ensembles flous des variables d'entrée et de sortie	25
3.1.2 Agrégation de la sortie des règles pour les individus 1 et 2	28
4.1.1 Estimation par l'assureur du nombre de jours d'usage réel du véhicule d'un assuré (exemple)	36
4.2.1 Sous-ensemble A représentant les bons conducteurs et les fonctions de distribution supérieure (F^*) et inférieure (F_*)	41
5.3.1 Définition des frontières du système bonus-malus selon deux approches	50

Liste des tableaux

3.1.1 Construction des nombres flous des variables d'entrée et de sortie	24
3.1.2 Fuzzification des variables Annual pct.driven et Avgdays.week pour deux individus	25
3.1.3 Construction des règles floues	27
4.1.1 Coefficients du GLM du contrat classique	34
4.1.2 Coefficients du GLM du contrat télématique	35
4.1.3 Résultats des choix de l'assuré selon différents scénarios	37
4.2.1 Quelques valeurs utiles à la construction des fonctions F^* et F_* pour le sous-ensemble A	41
5.2.1 Résultats des différents GLM comprenant les dimensions de l'ACP	47
5.2.2 Présentation des cinq sous-scores de conduite liés à la sinistralité	48
5.2.3 Résultats des sous-scores par catégorie	49
5.3.1 Variation de la prime selon le système bonus/malus avec deux sous-scores	51

Introduction

La théorie des ensembles est une branche des mathématiques qui permet de déterminer si un objet dans l'univers fait partie d'un ensemble particulier (Enderton, 1977). Déterminer si le résultat d'un lancer de dé fait partie d'un certain ensemble est évident. Cependant, l'aléatoire n'est pas la seule source d'incertitude présente dans le monde et si les limites d'un ensemble ne sont pas clairement définies, il est parfois difficile de classer un objet. Plusieurs causes peuvent amener de l'imprécision telles que la subjectivité de l'intervenant ou l'imprécision du langage (Wierman, 2010). La classification d'un conducteur illustre ce type d'incertitude. Quels critères déterminent si une personne est plutôt dangereuse, calme ou détient un comportement modéré au volant ? Tout dépend de la perspective du questionneur. Classer ce conducteur dans un ensemble bien défini est un acte quelque peu difficile. C'est en raison de ce problème que Zadeh (1965) a développé la théorie des ensembles flous qui tient compte de l'imprécision relative à certaines classes d'objets et permet l'adhésion graduelle d'un objet à un ensemble.

Dans le monde de l'assurance, le caractère flou est à privilégier pour modéliser l'imprécision et l'incertitude de certaines variables. La logique floue, qui est une extension de la logique classique, fournit un cadre conceptuel puissant dans la résolution des problèmes d'incertitude et d'imprécision linguistique (Yager et Zadeh, 1992). Ce mémoire a pour principal objectif la conception d'un produit télématique d'assurance automobile qui tient compte de ce caractère flou. Il est organisé comme suit :

- Le chapitre 1 définit la tarification de l'assurance automobile traditionnelle et défend l'apport que peuvent fournir les données télématiques à celle-ci.
- Le chapitre 2 évoque la théorie des ensembles flous et les nombreux concepts qu'elle détient tels que les nombres flous et le principe fondamental d'extension de Zadeh.
- Le chapitre 3 décrit la logique floue et la construction de systèmes d'inférence floue. Ces derniers étant particulièrement utiles pour la création d'un score flou de conduite.
- Le chapitre 4 parcourt les définitions principales de l'espérance floue. Des applications dans le domaine de l'assurance automobile sont évoquées.
- Le chapitre 5 conçoit un produit télématique par un système bonus-malus en utilisant la théorie des ensembles flous. Plusieurs sous-scores de conduite sont construits et utilisés afin de modifier la tarification de l'assurance automobile.

Chapitre 1

La tarification de l'assurance automobile et l'apport des données télématiques

1.1 Méthode classique de tarification de type GLM

1.1.1 Théorie du modèle

Les modèles linéaires généralisés (GLM) sont très largement utilisés en assurance dommages afin de tarifer une prime. Ils permettent d'expliquer le lien entre une variable à expliquer Y et un ensemble de variables explicatives X_1, \dots, X_p , au moyen d'une fonction de lien. L'utilisation du GLM en assurance est expliquée par de nombreux auteurs tels que [Denuit et Charpentier \(2005\)](#), [de Jong et Heller \(2008\)](#) et [Ohlsson et Johansson \(2010\)](#). En assurance automobile, la variable à modéliser est liée aux sinistres et peut, par exemple, se référer à la fréquence ou à la sévérité de ceux-ci. L'objectif est de déterminer la loi de probabilité de la variable réponse Y . La particularité de ce modèle est que cette dernière possède une densité de probabilité, notée $f_{\theta, \phi}$, appartenant à la famille exponentielle présente à l'équation (1.1.1) ([Denuit et al., 2019](#)).

$$f_{\theta, \phi}(y) = \exp\left\{\frac{y\theta - a(\theta)}{\phi/\nu}\right\} c\left(y, \frac{\phi}{\nu}\right), \quad y \in (-\infty, +\infty) \quad (1.1.1)$$

Où $a(\cdot)$ est une fonction convexe monotone qui dépend de θ , $c(\cdot)$ est une fonction définie sur \mathbb{R}^2 , $\theta \in \mathbb{R}$ est le paramètre canonique, ν est une constante positive connue appelée le poids et enfin $\phi \in \mathbb{R}_+$ est le paramètre de dispersion.

La régression linéaire suppose que toutes les données suivent une loi gaussienne, tandis que le GLM élargit ce modèle aux distributions exponentielles. De nombreuses lois classiques font partie de cette famille telles que la loi de Poisson, la loi Gamma ou la loi Normale. Étant donné que le vecteur X représente l'ensemble des variables explicatives, la meilleure prédiction de Y est la fonction de régression $E[Y|X]$. Contrairement au modèle linéaire classique où la moyenne n'est qu'une fonction linéaire des variables exploratoires,

c'est-à-dire telle que $E_\beta[Y|X] = X^T\beta$, le modèle linéaire généralisé insère cette partie linéaire ($X^T\beta$) dans la régression au moyen d'une fonction de lien inversible. Cette fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ est présente à l'équation (1.1.2).

$$g(E[Y|X]) = \eta = \beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p = X^T\beta \quad (1.1.2)$$

Où Y est la réponse, $X_1 \dots X_p$ sont les variables explicatives et β_0, \dots, β_p sont les coefficients de la régression. η est appelé le prédicteur linéaire. Les valeurs prédites par le modèle, $E[Y|X]$, sont donc obtenues par l'équation (1.1.3).

$$E[Y|X] = g^{-1}(\beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p) = g^{-1}(X^T\beta) \quad (1.1.3)$$

La fonction de lien $g(\cdot)$ dépend de la loi utilisée. Si la réponse Y admet une densité appartenant à la famille exponentielle, sa moyenne et sa variance sont définies par l'équation (1.1.4). Dès lors, θ influe à la fois l'espérance et la variance, tandis que le paramètre de dispersion ϕ influe uniquement la variance.

$$E[Y] = \mu = a'(\theta) \text{ et } V[Y] = a''(\theta) \frac{\phi}{\nu} \quad (1.1.4)$$

1.1.2 Hypothèses de base du modèle

Le GLM possède trois hypothèses importantes à garantir pour la validation des modèles statistiques (Ohlsson et Johansson, 2010).

Hypothèse 1 : indépendance des observations. Soit n polices différentes. Si Y_i désigne la réponse de l'observation i , alors Y_1, \dots, Y_n sont indépendantes. Bien que cette hypothèse soit indispensable, il arrive qu'elle ne soit pas respectée dans le domaine des assurances. Par exemple, pour une assurance automobile, une catastrophe naturelle causant de nombreux sinistres ne respecte pas cette hypothèse. Pour ce cas de figure, le GLM ne peut être utilisé. Dans ce mémoire, tout type d'évènement de ce genre est écarté et les observations sont considérées comme indépendantes.

Hypothèse 2 : indépendance temporelle. Soit n intervalles de temps disjoints. Si Y_i désigne la réponse dans l'intervalle de temps i , alors Y_1, \dots, Y_n sont indépendantes.

Hypothèse 3 : homogénéité. Soit deux polices possédant une prime identique et la même exposition. Si Y_i désigne la réponse de la police i , alors Y_1 et Y_2 ont la même distribution de probabilité.

1.1.3 Le cas des données de comptage

La régression de Poisson permet de modéliser la fréquence de sinistres. Ainsi, le nombre de sinistres est la variable Y à modéliser suivant un ensemble de variables explicatives X_1, \dots, X_i définissant chaque assuré. La variable Y suit une loi de Poisson de paramètre $\lambda > 0$ telle que sa densité de probabilité, notée $f(y)$, est définie par l'équation (1.1.5).

$$f(y) = \exp(-\lambda) \frac{\lambda^y}{y!} = \exp(y \ln(\lambda) - \lambda - \ln(y!)), y \in N. \quad (1.1.5)$$

Avec $\theta = \ln(\lambda)$, $a(\theta) = \lambda = \exp(\theta)$, $\phi = 1$ et $c(y, \phi) = -\ln(y!)$.

La fonction de lien canonique $g(\mu)$ est alors égale à $a'(\mu)$, c'est-à-dire à $\ln(\mu)$. Dès lors, $X^T \beta = \log(E[Y])$ et $E[Y] = g^{-1}(\mu) = \exp(X^T \beta)$. Le modèle devient alors multiplicatif.

1.1.4 Prise en compte de l'exposition

Pour tarifier un contrat, il est nécessaire de prédire le nombre de sinistres qui surviendront en moyenne l'année suivante. Pour ce faire, le nombre de sinistres survenus sur une police d'assurance est observé. Cependant, si toutes les polices n'ont pas été observées une année complète, il y a lieu de pondérer la fréquence de sinistres par l'exposition. Le processus de Poisson détient la propriété multiplicative, dans le sens où une police observée 1 an aura, en moyenne, 4 fois plus de sinistres qu'une police d'assurance observée 3 mois. Dès lors, l'équation (1.1.3) devient l'équation (1.1.6) si l'exposition est prise en compte.

$$E[Y|X] = \exp\{X^T \beta + \log(e)\} \quad (1.1.6)$$

Où e désigne l'exposition, en années.

1.2 L'apport des données télématiques dans l'assurance automobile

1.2.1 Le questionnement de l'assurance automobile traditionnelle

Dans toute assurance, l'analyse du risque réel est un enjeu primordial. Elle permet d'attribuer à chaque assuré la prime optimale nécessaire au bon fonctionnement du système. Toute compagnie d'assurance se doit de rester compétitive sur le marché en construisant le meilleur modèle. Celui-ci dépend bien entendu des prédictions réalisées grâce à l'information récoltée auprès de ses assurés. La mutualisation des risques entre des assurés présentant des caractéristiques communes est un mécanisme indispensable à l'assurance. Pour tout assureur, il est essentiel de segmenter son portefeuille en regroupant les risques homogènes et en leur attribuant une prime similaire. La prime est donc différente d'un groupe d'assurés à un autre. Traditionnellement en assurance automobile, des variables classiques telles que l'âge, l'état civil ou le type de véhicule utilisé sont employées afin d'établir cette segmentation.

Bien que cette information permette la segmentation, elle ne tient pas compte du comportement au volant de chaque assuré, ce qui peut conduire à de l'injustice. Un exemple très caricatural est celui de deux jeunes conducteurs âgés de 23 ans et possédant tous les

deux le même véhicule. Ces deux jeunes sont chez le même assureur et celui-ci risque fortement, toutes choses égales par ailleurs, de leur demander la même prime. Celle-ci sera par ailleurs certainement plus élevée que la moyenne de celles des autres assurés, vu leur très jeune âge (Bressand, 1993). Pourtant, il est probable qu'un des deux conducteurs conduise plus prudemment que l'autre (en faisant attention à sa vitesse et à ses virages). Malgré ce fait, ils paieront tout de même la même prime sur base de leurs variables traditionnelles.

Dans une vision plus économique, segmenter de manière optimale son portefeuille peut permettre à l'assureur d'obtenir un avantage concurrentiel. Dès lors, il est nécessaire pour ce dernier de ne pas manquer le tournant technologique lui permettant une meilleure segmentation par rapport à ses concurrents. C'est également un moyen d'éviter le phénomène d'anti-sélection. En effet, un avantage compétitif l'étant tout autant pour l'assureur que pour ses concurrents, celui-ci se pénalisera s'il n'en tire pas profit. Afin d'obtenir l'information liée au comportement routier de chaque assuré et de pouvoir l'utiliser de manière à adapter la tarification de l'assurance automobile, une solution pour l'assureur serait le recours à l'utilisation de la télématique.

1.2.2 L'utilisation des données télématiques

La télématique réfère à l'utilisation des appareils de télécommunication et des nouvelles technologies afin d'établir un contrôle à distance grâce au recueil de diverses données. De nombreuses applications de cette technologie apparaissent dans divers domaines tels qu'en journalisme, en système de santé ou dans le domaine privé. En ce qui concerne le secteur des assurances, celui-ci n'utilise pas la télématique à grande échelle. Cependant, cette technologie attire beaucoup d'attention dans le secteur automobile, où certaines données relatives au comportement routier sont récoltées à distance afin de modifier la tarification des assurances. C'est ce que l'on appelle l'assurance basée sur l'utilisation (UBI). Avec l'accord des assurés, certains véhicules peuvent être équipés d'un appareil télématique permettant de collecter et d'envoyer à l'assureur de nombreuses données concernant le conducteur. Ce dernier peut donc étendre son analyse actuarielle du risque automobile quant au lieu (type de route, circulation), au temps (kilométrage), au moment (l'heure, le jour de la semaine) et aux habitudes de conduite (vitesse, accélération, freinage, virage). Ces nouvelles variables permettent de mieux connaître l'exposition réelle des assurés sur la route.

Comme outil capable de percevoir ces données, des boîtes télématiques peuvent être installées dans les voitures individuelles. Ces boîtes sont munies de capteurs multiples, tels qu'un GPS et un géomètre, capables de détecter de nombreuses manœuvres diverses de conduites. Malheureusement, ces boîtes nécessitent un coût d'installation et d'entretien très élevé. De plus, le conducteur ne reçoit aucun feed-back pouvant améliorer son comportement au volant. Le déploiement est quant à lui très limité également. Dès lors, de nombreuses études se sont penchées sur l'utilisation des applications mobiles pour téléphone cellulaire pour le suivi des conducteurs (Handel et al., 2014; Araujo et al., 2012; Castignani et al., 2013b). Selon ces études, l'utilisation du smartphone est une technique

plus fiable et dispose de tous les capteurs nécessaires. Toutefois, l'évolution des technologies permet aux assureurs d'accroître leur panoplie d'outils de mesure, comme l'illustre le recours aux voitures connectées (Leroy, 2014).

Certains auteurs, dont Butler et al. (1988), relèvent l'inefficacité et l'injustice des primes forfaitaires ne tenant pas compte du nombre de kilomètres parcourus et de la manière de conduire des assurés. Actuellement, c'est pourtant la tendance de tarification des compagnies d'assurance automobile du monde entier. Grâce au recueil d'information concernant les habitudes de conduite que peuvent obtenir les assureurs grâce à la télématique, il est désormais possible pour eux d'améliorer l'élaboration du profil de risque de chaque conducteur. Avec ce système, le mécanisme de tarification ne se limitera donc plus aux données traditionnelles connues de tous telles que le sexe, l'âge ou l'état civil, mais tiendra compte de mesures réelles dynamiques. De plus, en améliorant leurs prévisions, les assureurs peuvent par conséquent donner aux conducteurs plus prudents la possibilité d'obtenir des primes plus abordables.

Une telle utilisation de la télématique peut être très avantageuse pour les jeunes conducteurs qui ont la possibilité d'obtenir des réductions de primes suivant leur comportement de conduite, dans le cas où ces derniers répondent aux exigences particulières d'un assureur (So et al., 2021a). Dans un système plus classique, les jeunes conducteurs ont généralement plus de chance d'avoir des primes d'assurance plus élevées que le reste de la population, car ceux-ci sont statistiquement plus à risque (Le Breton, 2001). La télématique peut également avoir un effet positif sur la société en général en encourageant un meilleur comportement au volant. Les consommateurs sont maintenant en mesure de contrôler le coût de leur assurance en améliorant leurs habitudes de conduite ou en réduisant la fréquence de leurs déplacements. La société en bénéficie aussi inconsciemment, car cela peut améliorer la sécurité routière et diminuer les accidents. Comme le souligne Zantema et al. (2008), innover la politique d'assurance peut avoir un impact significatif sur la sécurité. Différencier les primes en fonction du comportement sur les routes peut se montrer efficace.

Les progrès technologiques ont donné lieu à diverses variations de l'UBI. De nombreuses offres variées d'assurance automobile ont vu le jour, telles que "Pay as you Drive" (PAYD) qui se base sur l'exposition totale du conducteur en tenant compte du nombre de kilomètres parcourus, mais aussi du réseau routier utilisé afin de fixer une prime. Le modèle "Pay how you Drive" (PHYD) se base quant à lui sur le comportement de conduite de l'assuré (vitesse, accélérations brutales, virages brusques, freinages) (Tselentis et al., 2016). Il existe une réelle volonté croissante de rassembler de l'information sur le comportement des conducteurs afin de moduler le prix des polices d'assurance en fonction de leur niveau de risque.

1.2.3 Comportements de conduite liés aux accidents

Quelques études ont analysé les comportements dangereux de conduite qui peuvent amener à provoquer plus d'accidents. Par exemple, selon [Schöner \(2019\)](#), un comportement plus dangereux sur la route peut être expliqué par une action exceptionnellement brusque ou rapide, réalisée afin de rester sur la voirie ou de s'éloigner suffisamment des autres véhicules. Le danger routier peut également s'éveiller si les marges de sécurité routière (spatiales et temporelles) nécessaires sont réduites à cause des autres usagers de la route de telle sorte que les actions préventives deviennent compliquées en cas de risque élevé ([Shashua et al., 2018](#)). Comme le relève également [Lerner et al. \(2010\)](#), la conduite agressive, qui comprend des manœuvres telles qu'un changement soudain de la vitesse du véhicule, un changement rapide de la position latérale du véhicule, une accélération ou décélération brutale, est une cause majeure d'accidents de la route. Tous ces comportements sont des informations liées au risque routier que l'assureur peut récolter en utilisant la télématique.

1.2.4 Produits télématiques existants sur le marché de l'assurance automobile belge

Les produits d'assurance automobile ayant recours à la télématique se font encore rares en Belgique. Il est quand même intéressant de souligner la mise sur le marché de quelques produits tels que "We cover" de P&V, "driveXperience" d'Axa ou encore "Safety and More" (SAM) d'AG Insurance. L'assurance automobile "We cover" s'adresse plus particulièrement aux jeunes conducteurs de moins de 26 ans. Grâce aux données télématiques récoltées, la prime de chaque assuré peut varier selon le moment auquel le conducteur utilise son véhicule et le nombre de kilomètres parcourus. L'application offre également la possibilité au conducteur d'avoir une vue générale sur sa manière de conduire et lui permet ainsi de s'améliorer. Le produit "driveXperience" d'AXA attribue à chaque conducteur un score de conduite qui leur permet de réduire sa prime.

1.3 Intégration de la télématique au sein du GLM

1.3.1 Score de conduite construit grâce à la télématique

Il existe une littérature grandissante sur la télématique dans le domaine actuariel. Les prévisions de fréquence de sinistres, la classification des risques et les évaluations de primes sont, selon de nombreux travaux de recherche, plus précises grâce à la télématique. Pour tarifier un produit d'assurance automobile en intégrant des données télématiques, une des possibilités est l'ajout d'une variable au GLM. Cette variable peut représenter un score de conduite construit sur base de ces données qui dépendra donc de l'information liée au comportement routier de chaque assuré. Le risque de chacun sera quantifié et le modèle en sera par conséquent impacté. Un des objectifs de ce mémoire est la construction d'un tel

score.

1.3.2 Changements de comportement et gamification de l'assurance

La grande différence entre des variables classiques et télématiques est que l'assuré est capable de modifier ces dernières par son comportement. Par exemple, si un conducteur détient une vue sur son score de conduite (voir sous-section 1.3.1), ce dernier peut vouloir l'améliorer en changeant son comportement au volant. L'assureur peut également lancer des défis personnalisés à ses assurés pour améliorer leur conduite en leur offrant une récompense en cas de succès (diminution de prime). Cet aspect est fondé sur le principe de gamification et tend à rendre ludiques certaines actions parfois perçues comme ennuyantes. Ce service est donc destiné à améliorer la sécurité routière en récompensant les clients qui conduisent prudemment.

Il faut toutefois garder en tête que l'assuré a beaucoup moins de prise sur la sinistralité. Malgré le fait qu'il peut rendre sa conduite beaucoup plus souple ou moins agressive, le lien entre son comportement au volant et la sinistralité n'est pas évident. Un assureur peut suspecter que par exemple, au plus la vitesse d'un conducteur faisant un accident est élevée, au plus les dégâts et dommages seront conséquents. Cependant, aucun ouvrage ne démontre aujourd'hui ce lien comme étant direct et certain.

1.3.3 La caractéristique floue des données télématiques

Le comportement humain, notamment celui derrière le volant, comporte différents niveaux d'incertitude et le fait de stéréotyper celui-ci à l'aide de méthodes traditionnelles peut entraîner une perte d'informations. Ne pas tenir compte de cette incertitude revient à ne pas tenir compte du contexte et cela peut fausser le modèle. Par ailleurs, les capteurs utilisés au sein des voitures collectant les données télématiques peuvent également contenir de l'incertitude. Afin que celles-ci soient prises en compte, il est intéressant d'ajouter du bruit. Ce bruit peut être modélisé par la théorie des ensembles flous présente au chapitre 2. La théorie des ensembles flous est une approche flexible qui permet de capturer l'imprécision et l'incertitude contenues dans les données télématiques. Par exemple, dans la logique traditionnelle, un conducteur peut être considéré comme risqué dès qu'il effectue au minimum 25 accélérations brutales pour un certain trajet. Cela signifie qu'un assuré qui en réalise 24 n'est pas considéré comme risqué. Pourtant, il est évident qu'entre 25 et 24 accélérations brutales, le degré de risque est censé être assez proche voir quasiment identique. Cette illustration souligne l'importance et l'ajout de valeur que peut offrir une telle théorie. Le chapitre 3 évoquera la logique floue qui s'apparente au raisonnement humain, celle-ci étant fondée sur la théorie des ensembles flous.

Chapitre 2

Théorie des ensembles flous

2.1 Définition d'un ensemble flou

La théorie classique des ensembles développée par Georg Cantor (1845-1918) ne regroupe que deux états : soit un élément appartient à un ensemble et sa fonction caractéristique détient la valeur 1, soit il n'y appartient pas et sa fonction détient donc la valeur 0 (Vallin, 2013). Ainsi, pour un élément donné, seules deux possibilités sont tolérées, appartenir ou ne pas appartenir à un sous-ensemble. Un mathématicien renommé, du nom de Zadeh (1965), a voulu s'écarter de cette logique booléenne en développant une théorie des ensembles flous qui permet de représenter mathématiquement l'imprécision relative au regroupement de certains éléments dans des classes d'objets. En introduisant la notion de degré d'appartenance, il a permis des graduations dans l'appartenance d'un élément à un sous-ensemble et autorise ainsi un élément à appartenir un peu, beaucoup ou fortement à un sous-ensemble.

Si X représente l'ensemble universel, c'est-à-dire la collection de tous les objets candidats désignés génériquement par x , alors un ensemble flou A dans X est modélisé par un ensemble de paires ordonnées tel que :

$$A = (x, \mu_A(x) | x \in X)$$

Où $\mu_A(x)$ est une fonction caractéristique (d'appartenance) dont l'étendue est l'intervalle unitaire $U = [0, 1]$ telle que définie par l'équation (2.1.1).

$$\mu_A(x) = \begin{cases} 0 & \text{si } x \notin A \\ 0 < \mu(x) < 1 & \text{si } x \text{ a un degré d'appartenance à } A \\ 1 & \text{si } x \in A \end{cases} \quad (2.1.1)$$

Les ensembles universels sont, par exemple, les nombres réels ou les nombres naturels. Une fonction d'appartenance est donc une courbe qui associe à chaque objet x présent dans X la valeur de $\mu_A(x)$ comprise dans l'intervalle $[0, 1]$, valeur qui détermine le degré d'appartenance de x à l'ensemble flou A . Plus cette valeur est proche de 1, plus le degré

d'adhésion à cet ensemble flou est élevé. À l'inverse, au plus cette valeur est proche de 0, au moins cet élément appartient à l'ensemble flou. Si A est un sous-ensemble classique, la courbe d'appartenance qui lui est associée ne peut donc prendre que deux valeurs extrêmes, à savoir 0 et 1 (voir équation (2.1.2)).

$$\mu_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon.} \end{cases} \quad (2.1.2)$$

Il est souvent préférable d'utiliser la même notation pour un ensemble et pour sa fonction caractéristique. Par conséquent, dans ce mémoire, l'étiquette A sera utilisée pour représenter à la fois l'ensemble flou et sa fonction d'appartenance et la notation devient :

$$A(x) : X \longrightarrow [0, 1]$$

La seule différence entre un ensemble traditionnel et un ensemble flou est l'image de leurs fonctions d'appartenance. Les éléments d'un ensemble classique détiennent une appartenance totale dans un ensemble (degré d'appartenance égale à 1), et une appartenance nulle aux autres (degré d'appartenance égale à 0), comme expliqué auparavant. Dans un ensemble flou, les appartenances partielles sont comprises entre $[0, 1]$, car ce genre d'ensemble décrit des classes dont les limites ne sont pas clairement définies, des classes pour lesquelles de l'incertitude est présente. Un ensemble classique sera nommé "ensemble crisp" chaque fois qu'il sera nécessaire de le distinguer d'un ensemble flou. L'ensemble universel X est toujours un ensemble crisp.

La figure 2.1.1 permet d'illustrer la distinction entre un ensemble classique et un ensemble flou. Pour cet exemple, une variable d'entrée "nombre de virages brusques", qui compte pour chaque individu le nombre de virages brusques effectués par 1000 miles est inventée. Cet exemple vise à déterminer à partir de combien de virages brusques un conducteur peut être qualifié de "client à comportement hors norme". L'objectif est de créer un sous-ensemble qui contiendra tous les nombres de virages brusques correspondant à un comportement extrême. Dans un ensemble classique, par exemple, à partir de huit virages brusques sur dix, la situation est considérée comme extrême :

$$\text{Client à comportement hors norme} = \{\text{nombre de virages brusques} \geq 8\}$$

Par conséquent, le comportement d'un conducteur réalisant sept virages brusques n'est pas considéré comme anormal, bien qu'il ne lui restait seulement qu'un virage brusque à réaliser avant de le devenir. Cet ensemble classique est illustré à gauche de la figure 2.1.1. Cette appartenance complète ou nulle à un sous-ensemble n'est pas réaliste et cela souligne la nécessité de construire des sous-ensembles flous. L'équation (2.1.3) définit mathématiquement la fonction d'appartenance $A(x)$, qui détermine à quel degré un nombre x de virages brusques appartient au sous-ensemble "client à comportement hors norme". Elle est illustrée graphiquement à droite de la figure 2.1.1¹

1. Toutes les figures présentes dans ce mémoire ont été construites par l'auteur de ce mémoire.

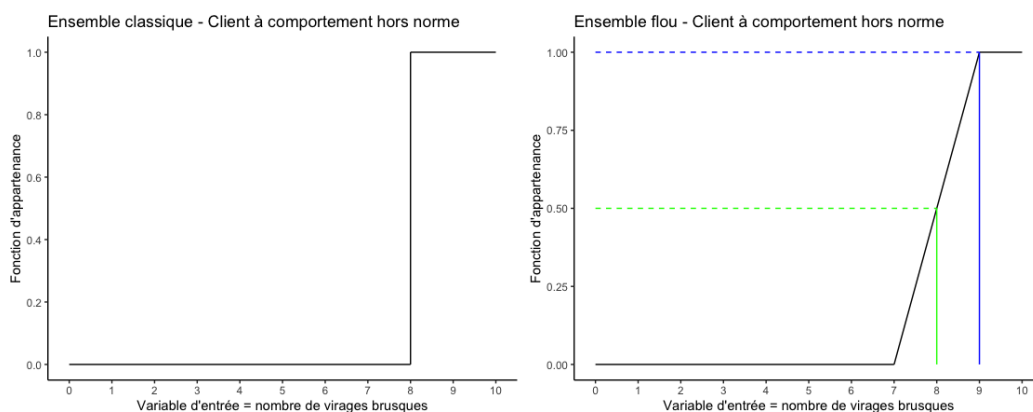


FIGURE 2.1.1 – Schéma descriptif d'un ensemble flou

$$A(x) = \begin{cases} 0 & \text{si } x \leq 7 \\ 0 < A(x) < 1 & \text{si } 7 < x < 9 \\ 1 & \text{si } x \geq 9 \end{cases} \quad (2.1.3)$$

Dès lors, à partir de neuf virages brusques, l'individu appartient totalement à l'ensemble client à comportement extrême (voir trait bleu sur le graphe) et il n'appartient pas du tout à ce sous-ensemble flou s'il réalise sept virages brusques ou moins. Si l'individu réalise huit virages brusques, il appartient à 50% à ce sous-ensemble (voir trait vert sur le graphe). Bien entendu, tous les nombres compris entre 7 et 9 détiennent un degré d'appartenance compris entre 0 et 1.

2.2 Opérations sur les ensembles flous

Dans cette section, les principales opérations applicables aux ensembles flous sont définies, telles que l'intersection floue, l'union floue et les compléments flous, qui sont respectivement représentés par les opérateurs "et", "ou" et "et pas" de la logique humaine (Wierman, 2010). La figure 2.2.1 représente graphiquement toutes ces opérations. Les notations utilisées postérieurement sont identiques à celles utilisées dans la théorie classique, car les ensembles "crisps" (nets) sont en réalité des cas particuliers des ensembles flous. En effet, ces deux types d'ensembles peuvent tous les deux être caractérisés par des fonctions d'appartenance. Comme un ensemble crisp détient des degrés d'appartenance uniquement de zéro et de un, valeurs comprises dans l'intervalle unitaire, cet ensemble peut être considéré comme un ensemble flou avec un ensemble d'images restreint. Le préfixe "flou" n'est donc pas indispensable pour ces opérations.

2.2.1 Sous-ensemble flou

Un ensemble flou A est un sous-ensemble d'un ensemble flou B si $A(x)$ est plus petit ou égal à $B(x)$ pour tout x dans X . Dans la théorie des ensembles flous, un sous-ensemble détient des éléments qui ont un plus petit degré d'appartenance. Dès lors, $A \subseteq B$ si et seulement si :

$$\forall x \in X : A(x) \leq B(x)$$

Pour qu'un ensemble flou A soit égal à un ensemble flou B , les valeurs d'appartenance de $A(x)$ et $B(x)$ doivent être égales pour tout x :

$$\forall x \in X : A(x) = B(x)$$

2.2.2 Intersection

L'intersection de deux ensembles flous A et B , notée $(A \cap B)(x)$ ou $(A \times B)$, est un ensemble flou dont la fonction d'appartenance est égale aux valeurs minimales des degrés d'appartenance de x dans A et B :

$$(A \cap B)(x) = \min[A(x), B(x)]$$

L'intersection correspond à la conjonction "et".

2.2.3 Union

L'union de deux ensembles flous A et B , notée $(A \cup B)(x)$ ou $(A + B)$, est un ensemble flou dont la fonction d'appartenance est égale aux valeurs maximales des degrés d'appartenance de x dans A et B :

$$(A \cup B)(x) = \max[A(x), B(x)]$$

L'union correspond à la conjonction "ou".

2.2.4 Complément

Le complément d'un ensemble flou A est un autre ensemble flou, noté A^c , dont la fonction d'appartenance est égale au degré opposé à celui que l'ensemble original A contenait :

$$A^c(x) = 1 - A(x)$$

Le complément correspond à la négation "et pas" du langage humain.

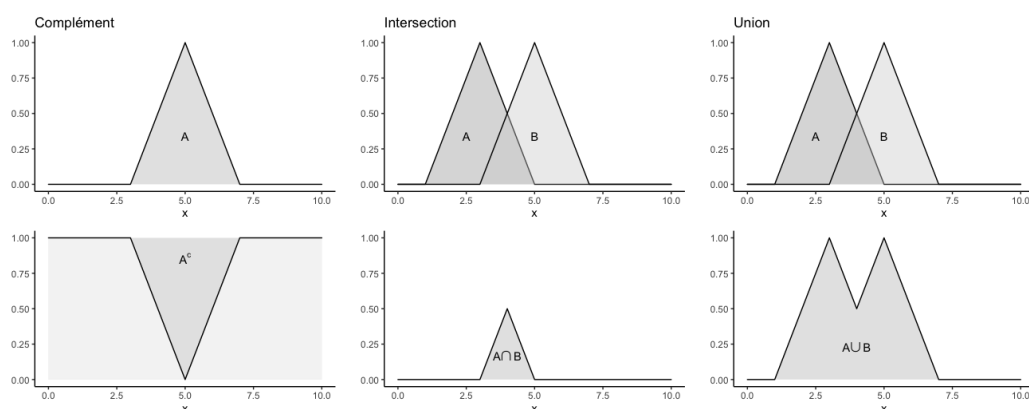


FIGURE 2.2.1 – Schéma des opérations sur les ensembles flous

2.3 Propriétés d'un ensemble flou

Un sous-ensemble flou A , défini par sa fonction d'appartenance $A(x)$, détient plusieurs propriétés importantes à définir dont la cardinalité scalaire, l'alpha coupe, le support, le noyau, la hauteur et le socle ([Wierman, 2010](#)).

2.3.1 La cardinalité scalaire

La cardinalité scalaire d'un ensemble flou fini A , notée $|A|$, est le nombre d'éléments compris dans A et peut se définir mathématiquement tel que :

$$|A| = \sum_{x \in X} A(x)$$

Comme certains éléments ne sont pas complètement dans A , la cardinalité scalaire n'est pas nécessairement un nombre entier.

2.3.2 L'alpha coupe (α -coupe)

L' α -coupe d'un ensemble flou A , notée A^α , est un sous-ensemble regroupant les éléments de A qui possèdent un certain degré minimal d'appartenance α :

$$A^\alpha = \{x \in X | A(x) \geq \alpha\}, \quad \forall \alpha \in [0, 1]$$

La coupe alpha d'un ensemble flou n'est pas un ensemble flou, mais bien un ensemble crisp. De plus, l' α -coupe fort, notée $A^{\alpha+}$, contient tous les éléments de A dont l'appartenance est strictement supérieure à alpha, tel que :

$$A^{\alpha+} = \{x \in X | A(x) > \alpha\}, \quad \forall \alpha \in [0, 1]$$

2.3.3 Le support

Le support d'un ensemble flou A de X , noté $S(A)$, est le sous-ensemble du domaine de tous les éléments qui lui appartiennent, c'est à dire le sous-ensemble du domaine qui a un degré d'appartenance positif, même très faible, à l'ensemble A . C'est l' α -coupe fort à zéro :

$$S(A) = A^{0+} = \{x \in X | A(x) > 0\}$$

2.3.4 Le noyau (ou pic, mode)

Le noyau d'un ensemble flou A de X , noté $C(A)$, est le sous-ensemble de tous les éléments de X qui appartiennent totalement à A . C'est la coupe alpha à 1 :

$$C(A) = A^1 = \{x \in X | A(x) = 1\}$$

2.3.5 La hauteur

La hauteur d'un ensemble flou A de X , notée $h(A)$, est le plus fort degré d'appartenance atteint par $A(x)$ sur le domaine de X :

$$h(A) = \max\{A(x) | x \in X\}$$

Où *max* désigne le maximum. En particulier, si $h(A) = 1$, l'ensemble flou A est normalisé. Il est toujours possible de normaliser un ensemble flou non vide en divisant sa fonction d'appartenance par sa hauteur. Le cœur d'un ensemble flou est l'ensemble de tous les éléments de X pour lesquels la valeur de la fonction d'appartenance est égale à la hauteur.

2.3.6 Le socle

Le socle d'un ensemble flou A de X , noté $p(A)$, est le plus faible degré d'appartenance atteint par $A(x)$ sur le domaine de X :

$$p(A) = \inf\{A(x) | x \in X\}$$

Où *inf* désigne le minimum, qui est la plus petite valeur que l'expression atteint lorsque α s'étend sur l'intervalle unitaire.

La figure 2.3.1 illustre schématiquement les différentes propriétés d'un ensemble flou A . Dans cet exemple, la cardinalité scalaire de l'ensemble flou A vaut deux, car le degré d'appartenance est nul pour les valeurs de x comprises entre zéro et trois et entre sept et dix, tandis qu'il vaut 0.5 quand x est égal à quatre ou six et 1 quand x est égal à cinq. Dès lors, la somme des degrés d'appartenance est bien égale à deux. Pour un alpha de 0.5, le schéma montre que l' α -coupe comprend les éléments de x étant égaux à 4, 5 et 6. Par ailleurs, le support de l'ensemble flou A est compris dans l'ensemble]3,7[. Le noyau vaut

cinq et la hauteur est égale à 1, ce qui signifie que le sous-ensemble flou A est normalisé. Pour finir, le socle vaut 0, car c'est le plus faible degré d'appartenance de cet ensemble flou.

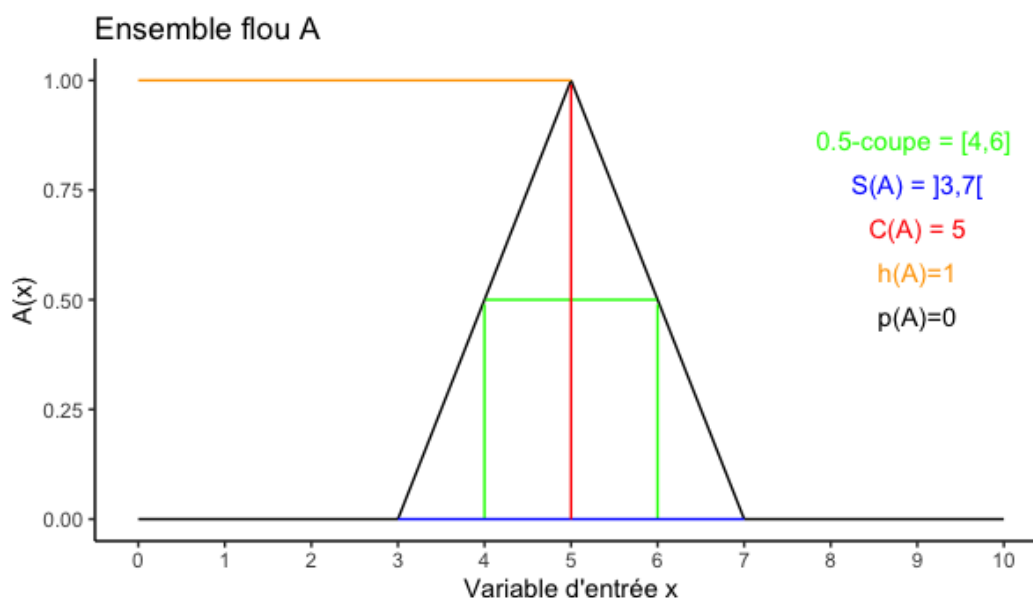


FIGURE 2.3.1 – Propriétés d'un ensemble flou

2.4 Nombres flous

Un nombre flou est une généralisation d'un nombre réel dans le sens où il représente une quantité dont la valeur est imprécise, plutôt qu'exacte comme c'est le cas pour les nombres "ordinaires" à valeurs uniques. Un nombre flou détient les propriétés floues. La quantité floue A avec une fonction d'appartenance $A(x)$ est un nombre flou si son domaine regroupe les nombres réels (\mathbb{R}) et s'il répond à quatre propriétés (Wierman, 2010) :

1. Normalité : un élément a le degré d'appartenance 1, $\exists x \in \mathbb{R}$ tel que $A(x) = 1$
2. Borné : le support de l'ensemble flou A ($S(A)$) est un ensemble borné de \mathbb{R}
3. Convexité : chaque α -coupe de A (A^α) est un intervalle fermé pour α positif.
4. La fonction caractéristique $A(x)$ est une fonction semi-continue supérieure.

Dès lors, les nombres flous sont des quantités floues dont les α -coupes sont des ensembles fermés et bornés : $A^\alpha = [\underline{a}, \bar{a}] = \{x \in X | A(x) \geq \alpha\}$, où :

$$\underline{a} = \min\{x \in X | A(x) \geq \alpha\} ; \bar{a} = \max\{x \in X | A(x) \geq \alpha\}$$

De nombreuses formes peuvent caractériser la courbe d'un nombre flou, telles qu'une forme de cloche (gaussienne), une forme triangulaire, une courbe trapézoïdale ou encore

pentagonale. Ce mémoire se limite à l'étude des courbes de distributions les plus populaires afin de construire les fonctions d'appartenance propres aux ensembles flous, à savoir les courbes trapézoïdales et triangulaires (voir figure 2.4.1).

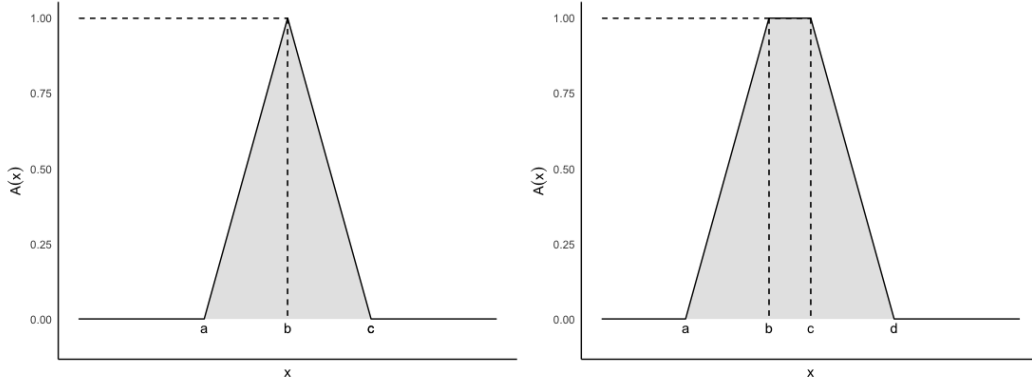


FIGURE 2.4.1 – Nombres flous d'une forme triangulaire (gauche) et d'une forme trapézoïdale (droite)

2.4.1 Nombres flous triangulaires

Un nombre flou de forme triangulaire, noté $Tr[]$, est généralement désigné par une triade (a,b,c) et est composé de deux segments. Mathématiquement, la forme d'un nombre flou triangulaire, noté $Tr[a, b, c]$ se trouve à l'équation (2.4.1).

$$Tr[a, b, c](x) = \begin{cases} 0 & \text{si } x < a \text{ ou } x > c \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ \frac{x-c}{b-c} & \text{si } b < x \leq c \end{cases} \quad (2.4.1)$$

$Tr[a, b, c](x)$ est la fonction de référence gauche pour $x \in [a, b]$ et la fonction de référence droite pour $x \in [b, c]$. La fonction de référence gauche est continue à droite, monotone et croissante, tandis que la fonction de référence droite est continue à gauche, monotone et décroissante. C'est un nombre flou qui monte en ligne droite du point $\langle a, 0 \rangle$ au point $\langle b, 1 \rangle$ puis redescend au point $\langle c, 0 \rangle$. L'alpha-coupe d'un nombre flou triangulaire, noté $Tr[a, b, c]^\alpha$ se présente sous la forme de l'équation (2.4.2).

$$Tr[a, b, c]^\alpha = [\underline{Tr[a, b, c]}, \overline{Tr[a, b, c]}] = [b - (1 - \alpha)(b - a), b + (1 - \alpha)(c - b)] \quad (2.4.2)$$

2.4.2 Nombres flous trapézoïdales

Un nombre flou de forme trapézoïdale ressemble à la forme triangulaire, mais reste à niveau à la hauteur de 1. Celui-ci, noté $Tp[a,b,c,d]$, avec $a \leq b \leq c \leq d$, est composé de trois segments : le premier allant du point $\langle a,0 \rangle$ au point $\langle b,1 \rangle$, le deuxième allant du

point $\langle b, 1 \rangle$ au point $\langle c, 1 \rangle$ et enfin le dernier allant du point $\langle c, 1 \rangle$ au point $\langle d, 0 \rangle$. Les intervalles entre a et b et entre c et d sont respectivement définis comme étant le côté gauche et le côté droit. La fonction d'appartenance d'un nombre flou trapézoïdal noté $Tp[a, b, c, d]$ se trouve à l'équation (2.4.3).

$$Tp[a, b, c, d](x) = \begin{cases} 0 & \text{si } x < a \text{ ou } x > c \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } b < x < c \\ \frac{x-d}{c-d} & \text{si } c \leq x \leq d \end{cases} \quad (2.4.3)$$

L'alpha-coupe d'un nombre flou trapézoïdal, notée $Tp[a, b, c, d]^\alpha$ se présente sous la forme de l'équation (2.4.4).

$$Tp[a, b, c, d]^\alpha = [\underline{Tp[a, b, c, d]}, \overline{Tp[a, b, c, d]}] = [b - (1 - \alpha)(b - a), c + (1 - \alpha)(d - c)] \quad (2.4.4)$$

Quand le côté gauche et le côté droit d'un nombre flou triangulaire ou trapézoïdal sont identiques, ces nombres flous sont symétriques.

2.5 L'arithmétique floue

2.5.1 Le principe d'extension de Zadeh

Le principe d'extension est un des outils les plus importants de la théorie des ensembles flous, car il permet de relier les ensembles flous à toutes les opérations, fonctions et outils des mathématiques classiques (Zimmermann, 2010).

Selon Zadeh (1965), le principe d'extension est défini comme suit. Soit un sous-ensemble flou A de l'univers classique X et soit une fonction f de X vers un univers Y . X et Y étant deux ensembles classiques. Alors, l'image de A par f , notée $f(A)$, est un nouveau sous-ensemble flou B de Y dont la fonction d'appartenance est définie par l'équation (2.5.1).

$$B(y) = \begin{cases} \sup_{x \in X} \{A(x) | y = f(x)\} & \text{si } f^{-1}(y) \neq \emptyset \\ 0 & \text{sinon} \end{cases} \quad (2.5.1)$$

La généralisation de ce principe à des fonctions de plusieurs variables est la suivante : soit X le produit cartésien de n univers $X_{i, i=1, \dots, n}$, $X = X_1 \times \dots \times X_n$ et soit A_1, \dots, A_n qui représentent n sous-ensembles flous définis dans les univers X_1, \dots, X_n , respectivement. f est une fonction qui fait correspondre X à un univers Y , $y = f(x_1, \dots, x_n) = f(x)$. $f^{-1}(y)$ désigne le sous-ensemble de X qui est mis en relation par la fonction f avec $y \in Y$, tel que $f^{-1}(y) = \{X | f(x) = y\}$. Dès lors, le principe d'extension permet de définir un ensemble flou, noté B , de domaine Y par l'équation (2.5.2). f^{-1} étant la fonction inverse de f .

$$B = \{(y, B(y)) | y = f(x_1, \dots, x_n), (x_1, \dots, x_n) \in X\}$$

Où :

$$B(y) = \begin{cases} \sup_{(x_1, \dots, x_n) \in f^{-1}(y)} \{ \min[A_1(x_1), \dots, A_n(x_n)] \} & \text{si } f^{-1}(y) \neq \emptyset \\ 0 & \text{sinon} \end{cases} \quad (2.5.2)$$

Ce principe d'extension permet de définir toute l'arithmétique des nombres flous. Par exemple, la somme et la multiplication de deux ensembles flous, notés A_1 et A_2 , définis dans les univers X_1 et X_2 respectivement, peuvent se définir par les deux équations suivantes :

$$(A_1 + A_2)(y) = \sup_{(x_1, x_2) \in A_1 \times A_2} \{ \min[A_1(x_1), A_2(x_2) | y = x_1 + x_2] \}$$

$$(A_1 \times A_2)(y) = \sup_{(x_1, x_2) \in A_1 \times A_2} \{ \min[A_1(x_1), A_2(x_2) | y = x_1 \times x_2] \}$$

Selon Colubi et al. (2007), l'alpha-coupe de la somme de deux nombres flous notés A_1 et A_2 , ainsi que l'alpha-coupe de la multiplication d'un nombre flou A_1 par un scalaire noté λ , peuvent respectivement être définies, $\forall \lambda \in \mathbb{R}$, par les équations suivantes :

$$(A_1 + A_2)^\alpha = \{x_1 + x_2 | x_1 \in A_1^\alpha, x_2 \in A_2^\alpha\} = [\inf\{A_1^\alpha\} + \inf\{A_2^\alpha\}, \sup\{A_1^\alpha\} + \sup\{A_2^\alpha\}]$$

$$(\lambda A_1)^\alpha = \{\lambda x_1 | x_1 \in A_1^\alpha\} = \begin{cases} [\lambda \inf\{A_1^\alpha\}, \lambda \sup\{A_1^\alpha\}] & \text{si } \lambda \geq 0 \\ [\lambda \sup\{A_1^\alpha\}, \lambda \inf\{A_1^\alpha\}] & \text{sinon} \end{cases}$$

2.5.2 Application du principe d'extension aux statistiques flous

Grâce au principe d'extension de Zadeh, Kruse et Meyer (1987) ont introduit la notion de paramètre flou d'une variable aléatoire floue. Soit un espace de probabilité (Ω, A, P) . Une variable aléatoire floue X peut être considérée comme une perception d'une variable aléatoire habituelle inconnue V , V étant l'originale de X . Soit χ l'ensemble de toutes les applications mesurables $V : \Omega \rightarrow \mathbb{R}$ par rapport à A . Un sous-ensemble de χ peut être défini par une fonction caractéristique ν qui correspond au degré d'appartenance qu'une variable aléatoire V soit l'originale de la variable aléatoire floue en question (voir équation (2.5.3)).

$$\nu(V) = \inf\{X(\omega)(V(\omega)) : \omega \in \Omega\} \quad (2.5.3)$$

Soit $\theta = \theta(V)$ un paramètre de la variable aléatoire réelle V . Pour la variable aléatoire floue X , il n'est pas possible d'observer directement le paramètre θ , mais seulement son image vague. Dès lors, la perception floue du paramètre θ est un ensemble flou $\Lambda(\theta)$ dont la fonction d'appartenance est donnée par l'équation (2.5.4).

$$\Lambda(\theta)(t) = \sup \left\{ \inf_{\omega \in \Omega} \{X(\omega)(V(\omega)) : V \in \chi, \theta(V) = t\} \right\} \quad t \in \mathbb{R} \quad (2.5.4)$$

Si $X = V$, il est évident que $\Lambda(\theta) = \theta$. Cette extension peut par exemple servir à l'élaboration de la définition d'une médiane floue d'une variable aléatoire X floue. Pour rappel, $\gamma \in \mathbb{R}$ est la médiane de la variable aléatoire V si elle satisfait aux inégalités suivantes :

$$D_V(\gamma^-) \leq 0.5 \leq D_V(\gamma)$$

Où D_V désigne la fonction de distribution de V . En utilisant l'équation (2.5.4), il est possible de généraliser le concept de médiane au contexte flou (Grzegorzewski, 2008). Si X est une variable aléatoire floue associée à l'espace de probabilité (Ω, A, P) , la médiane floue de X est le nombre flou $\Gamma(X)(t)$, tel que sa fonction d'appartenance, pour tout $t \in \mathbb{R}$, est définie par l'équation (2.5.5).

$$\Gamma(X)(t) = \sup \left\{ \inf_{\omega \in \Omega} \{X(\omega)(V(\omega)) : V \in \chi, D_V(\gamma^-) \leq 0.5 \leq D_V(\gamma)\} \right\}, \quad t \in \mathbb{R} \quad (2.5.5)$$

Cette extension sert de base aux définitions de toutes statistiques floues, telles que l'espérance ou la variance mathématique d'une variable aléatoire floue. De nombreux auteurs proposent une estimation de l'espérance ou de la variance floue. Le chapitre 4 est consacré à cette étude.

2.5.3 Distance entre deux nombres flous

De nombreuses définitions de la distance entre des ensembles flous existent dans la littérature. Parmi elles, celle de Bertoluzza et al. (1995) se montre particulièrement intéressante. Pour deux nombres flous A_1 et A_2 définis dans les univers X_1 et X_2 respectivement, la distance entre ceux-ci, notée D_W^ϕ est donnée par l'équation (2.5.6).

$$D_W^\phi(A_1, A_2) = \sqrt{\int_0^1 \int_0^1 [f_{A_1}(\alpha, \lambda) - f_{A_2}(\alpha, \lambda)]^2 dW(\lambda) d\phi(\alpha)} \quad (2.5.6)$$

Où $f_{A_1}(\alpha, \lambda) = \lambda \sup\{A_1^\alpha\} + (1 - \lambda) \inf\{A_1^\alpha\} = \lambda \bar{a}_1 + (1 - \lambda) \underline{a}_1$ (voir la propriété de l'alpha-coupe aux sous-sections 2.3.2 et 2.4.1). W et ϕ sont des mesures de pondération qui peuvent être identifiées à des mesures de probabilité sur l'espace mesurable $[0, 1]$. W est associé à une distribution non dégénérée, tandis que ϕ a une fonction de distribution strictement croissante sur $[0, 1]$. Gil et al. (2002) ont discuté de la manière dont le choix de W affecte la métrique D_W^ϕ . González-Rodríguez et al. (2006) ont quant à eux étudié le choix de ϕ .

Dans le chapitre 5, Puri et Ralescu (1986) ont utilisé cette notion afin de définir la variance d'une variable aléatoire floue (voir sous-section 4.5).

Chapitre 3

La logique floue

3.1 Les systèmes d'inférence floue

La logique floue est utilisée afin de prendre en compte l'imprécision et l'incertitude dans un processus de prises de décision. Elle s'emploie afin de synthétiser un protocole de contrôle linguistique d'un opérateur spécialisé. Deux principaux types de systèmes d'inférence floue existent, à savoir celui de Mamdani (1977) (Mamdani et Assilian, 1975) et celui de Takagi et Sugeno (1985) (TS). Ebrahim Mamdani a conçu son système flou afin d'étudier la possibilité d'une interaction humaine dans le contrôle d'une combinaison de machines à vapeur et de chaudières. Il a synthétisé un ensemble de règles floues obtenues auprès des personnes travaillant sur le système afin de procéder au contrôle. Le modèle TS se différencie du système de Mamdani, car les conséquences sont des fonctions des variables d'entrées. Ce modèle se base sur le premier modèle (Mamdani), car celui-ci est plus intuitif, globalement accepté et relativement bien adapté à la cognition humaine. Ce dernier aspect est très important, car le chapitre 5 se tournera vers la conception d'un produit télématique pour lequel les décisions du client prendront une grande importance.

Pour les opérations de classification, le système flou selon Mamdani consiste à mettre en relation un ensemble de variables d'entrée et de sortie à l'aide de fonctions d'appartenance et de règles linguistiques spécifiques. Afin de procéder à la construction d'un tel système, plusieurs étapes sont nécessaires. La première consiste à créer des ensembles flous et des fonctions d'appartenance (voir sous-section 3.1.1). S'en suivent l'étape de fuzzification (voir sous-section 3.1.2) et celle de la construction des règles floues (voir sous-section 3.1.3). La défuzzification est la dernière étape (voir sous-section 3.1.4). Comme illustration, un score flou de conduite allant de 0 à 100 sera calculé pour chaque assuré. Ce score dépendra du comportement routier et au plus il se rapprochera de 100, au plus l'assuré sera qualifié de calme au volant et représentera pour l'assureur un faible risque. Pour cet exemple, un sous-score de conduite relatif au temps passé sur la route est construit. Celui-ci dépendra de deux variables télématiques : "Annual pct.driven" et "Avgdays.week" (voir base de données annexe A.1 tableau A.1.1).

3.1.1 Création des ensembles flous et des fonctions d'appartenance

Tout d'abord, des ensembles flous sont créés pour chaque variable explicative et pour chaque variable de sortie, en divisant leurs domaines en plusieurs sous-régions nommées en termes linguistiques tels que "peu" ou "élevé". Il est possible de faire chevaucher un ensemble flou avec des ensembles flous voisins. De plus, il est commun qu'au plus il existe des niveaux intermédiaires, de sous couches linguistiques, au plus la classification sera précise. Cependant, au plus le nombre d'ensembles flous de chaque variable augmente, au plus le nombre de règles floues construites à l'étape suivante augmentera (voir sous-section 3.1.3) et au plus le système deviendra complexe. Afin de sélectionner le nombre idéal d'ensembles flous, ainsi que leurs limites pour chaque variable, il est nécessaire de procéder à des essais et de tester les erreurs. Après avoir créé les ensembles flous adéquats pour chaque variable, l'étape de fuzzification permet d'établir un lien entre des données réelles et ces sous-ensembles flous (voir sous-section 3.1.2).

Comme exemple, deux variables d'entrée sont utilisées, à savoir "Annual pct.driven" et "Avgdays.week". Annual pct.driven représente le pourcentage annuel du temps passé sur la route, tandis qu'Avgdays.week désigne le nombre moyen de jours d'utilisation du véhicule par semaine (voir annexe A.1). Le lien entre ces variables et la sinistralité est linéairement positif. Si Annual pct.driven ou Avgdays.week augmente, toutes choses égales par ailleurs, la fréquence de sinistres augmente également (voir annexe A.3.2 et A.3.1). Par conséquent, ces variables sont classées en trois sous-ensembles nommés "faible", "moyen" et "élevé". La variable de sortie "score de conduite" est créée et est classée en trois catégories, à savoir conducteur "agressif", "modéré" et/ou "calme". Au plus le score d'un assuré est élevé, au plus celui-ci sera qualifié de calme et donc de conducteur à faible risque. Les différents sous-ensembles flous de chaque variable sont de forme trapézoïdale et construits respectivement sur les domaines $X = [0, 1]$, $Y = [0, 7]$ et $Z = [0, 100]$. Ils sont représentés au tableau 3.1.1 et illustrés à la figure 3.1.1. Les limites de chaque sous-ensemble sont posées à la suite de l'analyse exploratoire.

Variable	Sous-ensembles flous	Français	Nombre flou trapézoïdal
Annual pct.driven	A_1	Faible	Tp[0,0,0.1,0.33]
	A_2	Moyen	Tp[0.17,0.4,0.6,0.83]
	A_3	Élevé	Tp[0.66,0.9,1,1]
Avgdays.week	B_1	Faible	Tp [0,0,1,3.5]
	B_2	Moyen	Tp[2.5,3.5,4.5,5.5]
	B_3	Élevé	Tp[4.5,6,7,7]
Score de conduite	C_1	Agressif	Tp [0,0,10,30]
	C_2	Modéré	Tp[20,40,60,80]
	C_3	Calme	Tp[70,90,100,100]

TABLE 3.1.1 – Construction des nombres flous des variables d'entrée et de sortie

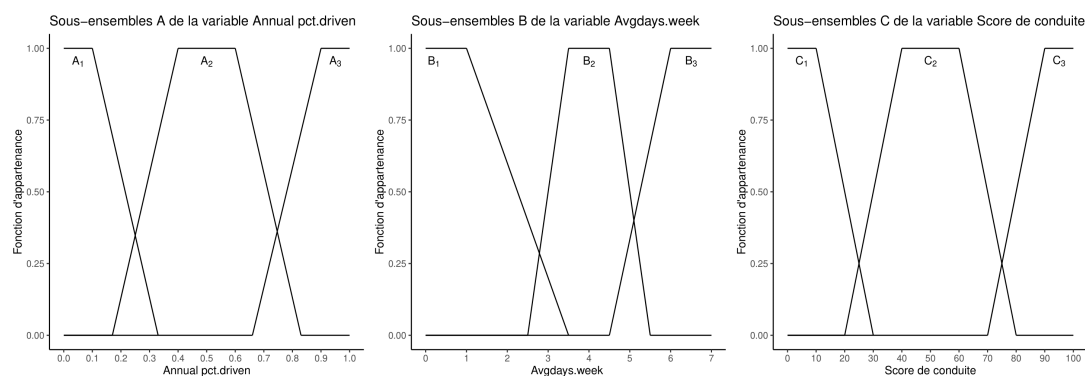


FIGURE 3.1.1 – Sous-ensembles flous des variables d'entrée et de sortie

3.1.2 Fuzzification

Grâce aux fonctions d'appartenance, l'étape de fuzzification permet d'établir le lien entre les données d'entrée réelles d'une base de données et le degré d'appartenance de celles-ci à chacun des ensembles flous. Pour chaque observation, les courbes d'appartenance de tous les ensembles flous sont calculées pour chaque variable et les sous-ensembles flous pour lesquels le degré d'appartenance n'est pas nul sont maintenus dans le système. Les combinaisons des ensembles flous restants sont ensuite évaluées par des règles floues afin d'obtenir des ensembles flous de sorties.

Pour exemple, deux individus présents dans la base de données ont été choisis. Il s'agit ici des données observées qui sont utilisées pour entrer dans le système de contrôle flou. Les valeurs des deux variables pour ces deux individus sont présentes au tableau 3.1.2. Ce tableau révèle que l'individu 1 appartient au sous-ensemble flou A_1 à 100% (Annual pct.driven= 6.6%) et aux sous-ensembles B_2 à 35% et B_3 à 43% (Avgdays.week= 5.15). L'individu 2 appartient totalement aux sous-ensembles A_2 et B_2 . Afin de connaître le degré d'appartenance des variables à chacun de ces sous-ensembles, il faut revenir à la définition d'un nombre flou trapézoïdal (voir équation (2.4.3)).

Annual pct.driven	individu 1	individu 2	Avgdays.week	individu 1	individu 2
x	0.066	0.551	x	5.15	4.18
$A_1(x)$	1	0	$B_1(x)$	0	0
$A_2(x)$	0	1	$B_2(x)$	0.35	1
$A_3(x)$	0	0	$B_3(x)$	0.43	0

TABLE 3.1.2 – Fuzzification des variables Annual pct.driven et Avgdays.week pour deux individus

3.1.3 Règles floues

L'étape de création des règles floues relie les différents ensembles flous d'entrées aux différents ensembles flous de sorties au moyen d'un raisonnement qualitatif. Ces règles floues sont des règles linguistiques de la forme :

"Si Entrée₁ est A_i et Entrée₂ est B_j , alors Sortie est C_k "

Où A_i , B_j et C_k représentent respectivement les sous-ensembles flous pour l'Entrée₁, l'Entrée₂ et la *Sortie*. Les valeurs des différentes entrées et sortie(s) sont les termes linguistiques choisis pour chaque ensemble à l'étape 1 (voir sous-section 3.1.1). La première partie de la règle se nomme l'antécédent (si ...), tandis que la fin (alors ...) est la conséquence. Le nombre de conditions, de même que les opérateurs flous, peuvent varier d'un système à l'autre en fonction d'un choix personnel. Dans cet exemple, l'opérateur "et" est utilisé (intersection de deux ensembles flous), mais il est également possible d'utiliser les termes "ou" (union de deux ensembles flous) et "et pas" (complément). La section 2.2 détaille les règles se référant à chaque opérateur flou.

Afin de construire ces règles floues, il est nécessaire de se baser sur l'expérience humaine. Il n'y a pas de bonnes ou mauvaises règles, mais il faut qu'il existe un lien logique. De plus, le poids de chaque règle influençant la sortie peut varier selon les besoins du système. Plusieurs propriétés doivent être mentionnées, dont celle des règles multiples (voir sous-section 3.1.3.1), des antécédents multiples (voir sous-section 3.1.3.2) et enfin celle de l'agrégation de la sortie des règles (voir sous-section 3.1.3.3).

3.1.3.1 Règles multiples

Si deux règles avec la même conséquence sont déclenchées pour une observation, la règle avec le degré d'appartenance maximum est utilisée pour l'inférence. Par exemple, si les deux règles suivantes se présentent :

Si Entrée₁ est A_1 , alors sortie est C_1

Si Entrée₂ est B_1 , alors sortie est C_1

La sortie C_1 est déclenchée pour cette observation avec un degré d'appartenance égale au maximum entre celui dont appartient l'entrée 1 au sous-ensemble A_1 et celui dont appartient l'entrée 2 au sous-ensemble B_1 .

3.1.3.2 Antécédents multiples

Si une règle floue comporte plusieurs antécédents, l'opérateur flou utilisé afin de les unir est décisif pour connaître le degré d'appartenance de la sortie. Si la conjonction qui les unit est le mot "et", le minimum des deux degrés d'appartenance est choisi. Au contraire, si la conjonction "ou" est utilisée pour relier deux antécédents, c'est le maximum des deux qui est ici retenu.

3.1.3.3 Agrégation de la sortie des règles

Cette étape permet d'unifier les sorties de toutes les règles. Pour chaque observation, un seul ensemble flou de sortie est créé à partir de toutes les règles. Pour ce faire, le contrôleur flou unit (voir sous-section 2.2.3) toutes les conclusions afin d'obtenir l'ensemble flou final. Il fusionne les sous-ensembles flous de sortie tronqués à l'aide de l'opérateur max pour produire l'ensemble flou résultant.

Dans l'exemple, des règles pour toutes les combinaisons possibles sont définies. Comme il y a deux variables qui ont chacune trois sous-ensembles flous ($i = 3, j = 3$), 9 règles (3×3) sont nécessaires pour le traitement de toutes les valeurs d'entrée comprises dans le domaine (voir tableau 3.1.3). Les trois premières règles présentes dans ce tableau sont logiques. Si un individu détient deux variables qui représentent un faible/moyen/élevé risque, alors il sera respectivement qualifié de conducteur calme/modéré/agressif. Pour les règles 4 et 5, si un individu détient une variable présentant un faible risque et une autre qui en présente un élevé, il sera qualifié de conducteur modéré. Pour finir avec les quatre dernières règles, comme aucune variable de sortie ne peut faire le compromis entre les deux variables, c'est la variable la plus significative en ce qui concerne la sinistralité qui prend le dessus. Dans cet exemple, la variable "Annual pct.driven" prend le dessus, car c'est celle qui est la plus significative. Pour le savoir, un GLM sur base de ces deux variables télématiques a été réalisé.

Numéro de la règle	Annual pct.driven	Et	Avgdays.week	Alors	Score de conduite
1	Faible (A_1)		Faible (B_1)		Calme (C_3)
2	Élevé (A_3)		Élevé (B_3)		Agressif (C_1)
3	Moyen (A_2)		Moyen (B_2)		Modéré (C_2)
4	Élevé (A_3)		Faible (B_1)		Modéré (C_2)
5	Faible (A_1)		Élevé (B_3)		Modéré (C_2)
6	Faible (A_1)		Moyen (B_2)		Calme (C_3)
7	Moyen (A_2)		Faible (B_1)		Modéré (C_2)
8	Élevé (A_3)		Moyen (B_2)		Agressif (C_1)
9	Moyen (A_2)		Élevé (B_3)		Modéré (C_2)

TABLE 3.1.3 – Construction des règles floues

Par conséquent, deux règles floues s'enclenchent pour l'individu 1, à savoir les règles numéro 5 et 6 :

Si Entrée₁ est A_1 à 100%, Et Entrée₂ est B_2 à 35%, alors sortie est C_3 à 35%

Si Entrée₁ est A_1 à 100%, Et Entrée₂ est B_3 à 43%, alors sortie est C_2 à 43%

Pour chaque règle, comme l'opérateur flou "et" est utilisé, le degré d'appartenance du sous-ensemble de sortie est égal au minimum de ceux des antécédents (voir sous-section 3.1.3.2). Comme conclusion, l'individu 1 appartient à 35% au sous-ensemble C_3 et

à 43% au sous-ensemble C_2 . Pour obtenir l'ensemble flou final C , ces deux sous-ensembles flous tronqués sont fusionnés à l'aide de l'opérateur max. Pour l'individu 2, seule la règle flou numéro 3 se déclenche et l'ensemble final C est égal à C_2 . L'ensemble flou final C pour chaque individu est présent à la figure 3.1.2.

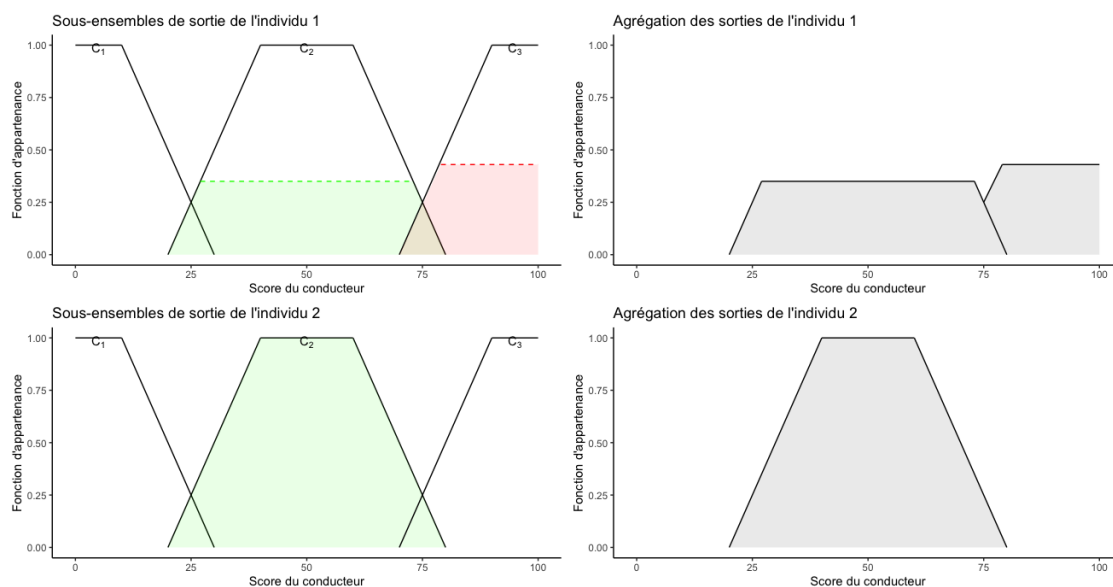


FIGURE 3.1.2 – Agrégation de la sortie des règles pour les individus 1 et 2

3.1.4 Défuzzification par la méthode du centre de gravité (COG)

La dernière étape de la construction du système flou consiste à produire un seul et unique résultat quantifiable de classification en sortie. La défuzzification est le mécanisme qui permet de convertir une valeur floue en une valeur nette afin d'obtenir une décision concrète. Plusieurs méthodes de défuzzification sont proposées dans la littérature. [Leekwijck et Kerre \(1999\)](#) distinguent trois groupes de techniques de défuzzification, à savoir les méthodes de maxima, les méthodes de distribution et les méthodes de zone. Chaque méthode présente des avantages et des inconvénients et le choix de l'une d'entre elles dépend de l'analyse voulue et du type de données ([De Runz et al., 2008](#)). Les méthodes de maxima considèrent que la valeur de défuzzification doit se trouver parmi les éléments du cœur de l'ensemble flou de sortie agrégé (voir notion de cœur à la sous-section 2.3.5). Parmi elles, la méthode "first of maximum" (FOM) désigne le plus petit élément du cœur ; la méthode "last of maximum" (LOM) prend le plus grand élément du cœur et enfin la méthode "middle of maximum" (MOM) calcule l'élément médian du cœur. Ces méthodes sont puissantes au niveau du calcul. Les méthodes de zone se basent quant à elles sur l'aire sous la courbe représentative de la fonction d'appartenance. Pour terminer, les méthodes de distribution convertissent la fonction d'appartenance en une fonction de densité de probabilité afin de calculer la valeur attendue. L'hypothèse de continuité est vérifiée dans ces méthodes.

Parmi ce dernier groupe, la méthode du centroïde reste la technique la plus populaire et une tactique courante dans les articles sur l'assurance est d'utiliser cette dernière. Ce mémoire se base exclusivement sur celle-ci. Des détails quant à la sélection et à l'utilisation d'une méthode de défuzzification sont fournis par [Runkler \(1997\)](#). Le choix de la méthode de défuzzification a un impact important sur le modèle et sur les décisions qui en découlent.

La méthode de défuzzification par le centre de gravité est une méthode générale qui calcule le centre de gravité de l'aire sous la fonction d'appartenance qui caractérise l'ensemble flou de sortie issu de l'agrégation des conclusions/conséquences. Si un ensemble flou se note A et que sa fonction d'appartenance est $A(x)$, alors son centre de gravité (CM) se calcule par l'équation (3.1.1).

$$CM = \frac{\sum_{x=x_{min}}^{x=x_{max}} A(x) \times x}{\sum_{x=x_{min}}^{x=x_{max}} A(x)} \quad (3.1.1)$$

Selon cette méthode, l'individu 1 détient un score de conduite de 60.5, tandis que celui de l'individu 2 est de 50. Par conséquent, selon ce score de conduite, l'individu 1 est considéré comme un assuré plus calme sur la route et moins dangereux.

3.2 Performance de la théorie

La performance d'un tel système repose principalement sur les choix pris par le contrôleur concernant la construction des ensembles flous d'entrée et de sortie, la construction des courbes d'appartenance et l'ajustement des règles floues liant les entrées et les sorties. Une analyse exploratoire sur une partie du jeu de données permet d'initier les paramètres. Il est intéressant de parcourir les avantages (voir sous-section 3.2.1) et les inconvénients (voir sous-section 3.2.2) d'un tel système.

3.2.1 Avantages

Le premier avantage de l'utilisation de la logique floue est que celle-ci passe des variables numériques traditionnelles à des variables linguistiques qui peuvent se rapprocher de la pensée humaine. Par exemple, elle peut permettre d'expliquer pédagogiquement aux assurés l'évolution d'un score d'un produit télématique. Cette méthode rend la compréhension plus facile et contourne l'élaboration d'équations analytiques complexes. De plus, un tel système peut se mettre en place de manière assez rapide, sans connaître les détails les plus minutieux du modèle. Un autre avantage est le coût du processus, qui est relativement faible. Pour finir, la plage d'adaptation du modèle est assez large, car il existe un panel de fonctions d'appartenance facile à mettre au point.

3.2.2 Inconvénients

Ce système possède cependant quelques inconvénients. Tout d'abord, les explications fournies par le modèle sont assez limitées et il est nécessaire de consulter des experts pour l'élaboration des règles floues. De plus, comme de nombreuses définitions des fonctions d'appartenance existent, il y a un risque de mal les définir ou de ne pas construire le modèle le plus optimal. En assurance automobile, le lien fort entre les entrées (ensembles flous, règles floues) et les conséquences (variation de primes) nécessite des études statistiques pour assurer l'équilibre du portefeuille.

3.3 L'utilisation de la logique floue en assurance

Cette section vise à souligner quelques applications des systèmes d'inférence floue qui existent dans le domaine de l'assurance. La logique floue s'est appliquée pour la première fois au domaine de l'assurance en 1982. Cette application visait à quantifier le caractère flou de la tarification de la souscription d'un contrat d'assurance vie individuel en utilisant un système d'inférence floue (de Wit, 1982). Par après, Shapiro (2007) a également utilisé les systèmes d'inférence floue dans plusieurs domaines liés à l'assurance tels que pour la tarification de l'assurance-vie et de l'assurance-maladie, la modélisation du processus de sélection dans l'assurance-maladie collective, l'évaluation des risques (y compris le risque d'accident du travail), la tarification de l'assurance-maladie collective à l'aide de données auxiliaires floues, des prévisions financières et la budgétisation des soins de santé nationaux.

Cette théorie a également été employée dans le but de fournir une définition flexible de l'assuré privilégié en assurance vie. Lemaire (1990) s'est concentré sur quelques variables, dont le cholestérol, la pression artérielle, le poids et le statut de fumeur. Hosler (1992) s'est également emparé de ce type de système afin de modéliser le processus de sélection dans l'assurance maladie collective. De plus, des auteurs ont montré comment un système d'inférence floue peut être utilisé dans la tarification du diabète sucré (Horgby et al., 1997). C'est un bon exemple de situations médicales où il est facile de parvenir à un consensus entre médecins sur le caractère léger, modéré ou grave d'une maladie ou d'un symptôme, mais où il n'existe normalement pas de méthode permettant de quantifier cette évaluation. En Belgique, Mosmans et al. (2002) développent des outils méthodologiques visant à aider à comprendre la consommation du budget des soins de santé. L'accent est mis sur le développement d'un système d'aide à la décision mettant en œuvre la méthodologie élaborée.

Plus récemment, des chercheurs ont mis en place un système d'inférence floue qui combine des données provenant de capteurs les plus courants intégrés dans les smartphones actuels afin d'évaluer le comportement de conduite global de chaque assuré (Castignani et al., 2013a,b). Toutes ces recherches montrent et illustrent le caractère unique que peut avoir l'assurance en tant que domaine d'application. Bien que de nombreuses autres applications existent dans ce domaine, il n'est cependant pas possible de toutes les parcourir.

Chapitre 4

Statistiques floues et application sur une variable télématique

Dans ce chapitre, l'objectif est d'adapter les définitions classiques de l'espérance et de la variance aux nombres flous, ces deux statistiques étant des notions importantes dans le domaine actuariel. Mathématiquement, la prime pure d'un contrat d'assurance est égale à l'espérance des pertes. Dans la littérature, de nombreux auteurs détiennent leur propre définition de l'espérance floue et il est intéressant d'en parcourir certaines. Tout d'abord, la définition de l'espérance floue établie par Gonzalez et de Campos sera parcourue à la section 4.1, suivie de celle construite par Dubois et Prade à la section 4.2. Pour ces deux définitions, une application dans le domaine de l'assurance automobile sera explorée. Afin de montrer la grande panoplie des outils flous, les définitions de l'espérance floue selon Kandel (voir section 4.3), dans un espace de crédibilité (voir section 4.4) et selon Puri et Ralescu (voir section 4.5) seront brièvement abordées. Dans la section consacrée à Puri et Ralescu, une définition de la variance floue, liée à la distance entre deux nombres flous calculée par Bertoluzza à l'équation (2.5.6) (voir chapitre 2 sous-section 2.5.3), sera également proposée.

4.1 Classification floue selon Gonzalez et De Campos

4.1.1 Théorie

En ce qui concerne la classification des nombres flous, deux approches assez différentes peuvent ressortir. D'une part, il y a un classement des nombres flous à l'aide d'un ensemble qualifié de "crisp" (de Campos Ibáñez et González Muñoz, 1989; Adamo, 1980). D'autre part, se trouve une approche qui construit une relation floue en donnant, pour chaque paire de nombres flous, un indice de comparaison. Quelques exemples tirés de cette approche sont les travaux de Dubois et Prade (1983) ou ceux de Delgado et al. (1988).

Dans cette sous-section, l'étude de l'approche subjective de classification réalisée par

de Campos Ibáñez et González Muñoz (1989) est analysée. Ces derniers utilisent la subjectivité du décideur afin d'obtenir un indice de classement. Tout d'abord, le décideur choisit un sous-ensemble Y compris dans l'intervalle unitaire, $Y \subseteq [0, 1]$, de sorte que les ensembles de niveaux qui y sont associés contiennent l'information principale de la quantité (celle-ci étant imprécise). Un poids représenté par une mesure de probabilité P est ensuite attribué aux différents éléments de Y . Le décideur détermine également la fonction paramétrique, notée $f_A^\lambda(\alpha)$, qui représente la position de chaque alpha-coupe (voir théorie à la sous-section 2.3.2) d'un nombre aléatoire flou A dans l'ensemble réel (\mathbb{R}). Les valeurs de cette fonction vont ensuite être intégrées par les poids représentant l'importance subjective de chaque ensemble de niveaux. Cette fonction, appelée indice moyen, peut donc être interprétée comme une moyenne pondérée des positions des ensembles de niveaux dans Y en utilisant P . Pour choisir $f_A^\lambda(\alpha)$, il est conseillé de prendre un point inclus dans chaque ensemble de niveaux A^α de A , tel qu'à l'équation (4.1.1).

$$f_A^\lambda : Y \rightarrow \mathbb{R}, \quad f_A^\lambda(\alpha) = \lambda \bar{a} + (1 - \lambda) \underline{a} \quad (4.1.1)$$

Où $\lambda \in [0, 1]$ et $A^\alpha = [\underline{a}, \bar{a}]$. Le paramètre λ reflète le degré d'optimisme ou de pessimisme du décideur. Ce paramètre va permettre au décideur de se rapprocher de la quantité qui lui semble la plus optimale.

La valeur moyenne d'un nombre aléatoire flou A , notée $V_P^\lambda(A)$, peut être définie par l'équation (4.1.2).

$$V_P^\lambda(A) = \int_Y f_A^\lambda(\alpha) dP(\alpha) \quad (4.1.2)$$

Où P est une mesure de probabilité sur Y . Grâce à cet indice, l'équation (4.1.3) permet de comparer deux nombres aléatoires flous A et B , présents dans un ensemble flou N .

$$\forall A, B \in N : A \leq B \Leftrightarrow V_P^\lambda(A) \leq V_P^\lambda(B) \quad (4.1.3)$$

De plus, A est indifférent de B si leurs indices moyens coïncident (voir équation (4.1.4)).

$$A \approx B \Leftrightarrow V_P^\lambda(A) = V_P^\lambda(B) \quad (4.1.4)$$

La proposition 1 montre que la valeur moyenne d'un nombre aléatoire flou A , notée $V_P^\lambda(A)$, est une fonction linéaire par rapport au paramètre λ .

Proposition 1. *Soit \tilde{R} l'ensemble des nombres flous, soit $A \in \tilde{R}$ un nombre aléatoire flou, soit $t, \mu \in [0, 1]$. Alors,*

$$V_P^{\lambda t + (1-\lambda)\mu}(A) = \lambda V_P^t(A) + (1 - \lambda) V_P^\mu(A), \forall \lambda \in [0, 1]$$

À la sous-section 4.2.2, un lien sera effectué entre la classification floue selon Gonzalez et de Campos et une autre théorie qui est celle de l'espérance floue selon Dubois et Prade (voir section 4.2). Pour ce faire, la mesure de Lebesgue, notée P_L , sera utilisée comme

mesure de probabilité. Lorsque la mesure de Lebesgue sur $Y = [0, 1]$ est utilisée, la valeur moyenne d'un nombre aléatoire flou A sera notée par $V_{P_L}^\lambda(A)$.

4.1.2 Application : choix de l'assuré concernant le type de contrat

Cet outil est intéressant dans le domaine de l'assurance automobile, car il peut permettre d'évaluer la position de l'assuré concernant le choix, ou non, d'acquérir un contrat télématique. Afin d'estimer ce choix, il y a lieu de comparer la prime pure d'un contrat traditionnel (voir sous-section 4.1.2.1) à celle provenant de l'exploitation des données télématiques (voir sous-section 4.1.2.2). La prime pure d'un contrat est le montant que la compagnie d'assurance doit demander afin de pouvoir indemniser tous les sinistres, sans perte ni profit. Selon la forte loi des grands nombres, celle-ci est l'espérance de la perte globale d'un certain portefeuille d'assurance et se calcule en multipliant la fréquence moyenne de sinistres avec le coût moyen de ceux-ci. Cependant, pour plus de simplicité et afin de se concentrer sur l'aspect télématique, seule la fréquence des sinistres est modélisée afin de prédire la prime. Par conséquent, si la fréquence de sinistres augmente avec un modèle, la prime pure associée augmente également, toute chose étant égale par ailleurs.

Afin d'analyser l'effet d'une variable télématique dans un contrat d'assurance automobile, la variable "Avgdays.week", qui représente pour chaque assuré le nombre moyen de jours passés sur la route par semaine, est utilisée comme exemple (voir variable "Avgdays.week" en annexe A.1 dans le tableau A.1.1). Comme cette variable est positivement corrélée avec la sinistralité (voir annexe A.3.1), en optant pour un contrat télématique, l'assuré aura probablement la possibilité de diminuer sa prime s'il utilise moins sa voiture. Au contraire, si ce dernier l'utilise énormément, sa prime peut bien entendu augmenter et son choix se tournera plutôt vers un contrat d'assurance classique. L'assuré aura donc le choix entre un contrat classique fixant une prime pure sur base de variables classiques telles que son âge ou le type de véhicule utilisé ou un contrat qui, en plus de ces mêmes variables classiques, utilisera la variable télématique "Avgdays.week" afin de calculer la prime pure du contrat. Ce contrat télématique permet donc à l'assuré de modifier sa prime en fonction de ses habitudes de conduite.

4.1.2.1 Prime pure d'un contrat traditionnel

La première étape consiste à calculer la prime pure d'un contrat traditionnel. Pour rappel, afin de simplifier le modèle, seule la modélisation de la fréquence de sinistres est prise en compte. Pour ce faire, un modèle linéaire généralisé (GLM) selon une loi de Poisson est effectué (voir théorie sous-section 1.1.1). Afin de raisonner par classes de risques, il est nécessaire de supposer que les variables explicatives sont des variables qualitatives. Dans un premier temps, les variables classiques continues de la base de données (voir annexe A.1) sont donc coupées en variables catégorielles. Cette étape facilite l'interprétation des résultats et permet de ne pas avoir besoin d'hypothèses de linéarité (méthode non paramétrique). Pour maintenir une certaine homogénéité dans la taille de chaque groupe, ces

coupages ont été réalisés au niveau des différents quartiles. Ensuite, un premier GLM est établi sur toutes les variables classiques. La variable "Duration", déterminant pour chaque assuré la durée de couverture d'assurance en jours, est l'offset du modèle. En effet, comme les couvertures d'assurance des polices n'ont pas toutes la même durée (variable comprise entre 22 et 366 jours), il est nécessaire de pondérer la fréquence de sinistres par l'exposition afin de prédire le nombre de sinistres qui surviendront l'année suivante (voir sous-section 1.1.4). Une sélection est ensuite réalisée pas à pas afin de sélectionner le sous-ensemble de variables qui construira le modèle le plus performant, c'est-à-dire celui qui aura l'erreur de prédiction la plus faible. Pour ce faire, la fonction `step()` est appelée dans R. Finalement, les coefficients présents au tableau 4.1.1 sont obtenus.

Variable	Coefficient	Variable (suite)	Coefficient
Intercept	$\beta_0 = -8.174$	Credit.score between 766 and 825	$\beta_8 = -0.414$
Insured age between 39 and 51 years	$\beta_1 = 0.113$	Credit.score between 825 and 856	$\beta_9 = -0.721$
Insured age between 51 and 63 years	$\beta_2 = 0.065$	Credit score more than 856	$\beta_{10} = -0.719$
Insured age more than 63 years old	$\beta_3 = -0.102$	Region Urban	$\beta_{11} = 0.189$
Car age more than 5 years old	$\beta_4 = -0.467$	More than 7456.452 annual miles drive	$\beta_{12} = 0.176$
Car.use Commute	$\beta_5 = -0.225$	Between 15 and 29 years no claims	$\beta_{13} = -0.268$
Car.use Farmer	$\beta_6 = -0.858$	Between 29 and 41 years no claims	$\beta_{14} = -0.263$
Car.use Private	$\beta_7 = -0.226$	More than 41 years no claims	$\beta_{15} = -0.108$

TABLE 4.1.1 – Coefficients du GLM du contrat classique

Soit un homme assuré de 71 ans qui a une duration de 183 jours et qui dispose d'une voiture privée de 6 ans d'ancienneté. Son score de crédit est égal à 842 et il vient d'une région urbaine. Annuellement, il a roulé 6213.710 miles et cela fait 43 ans qu'il n'a pas déclaré d'accident. La fonction `predict()` dans R prévoit que la prime de cet assuré se calculerait sur base d'une fréquence de sinistres égale à 0.01229 pour ce contrat. Cette fonction réfère à l'équation (1.1.6) de la sous-section 1.1.4.

4.1.2.2 Prime pure d'un contrat télématique

La seconde étape consiste à calculer la prime pure d'un contrat télématique en ajoutant la variable télématique "Avgdays.week" aux variables classiques. Cette variable dépend des habitudes réelles de conduite de chaque assuré et peut être dérivée en trois sous-ensembles flous A, B et C. Le sous-ensemble A est dédié au faible nombre de jours passés sur la route, B à un nombre moyen et C à un nombre élevé. Comme il s'agit de la même variable télématique que celle utilisée au chapitre 3, les mêmes choix concernant la construction de ces sous-ensembles flous ont été pris (voir sous-section 3.1.1). Soit un assureur qui souhaite proposer un produit télématique à son assuré et demande à ce dernier de choisir, parmi les trois catégories suivantes, celle qu'il pense lui correspondre :

- La catégorie 1 qui regroupe les conducteurs prenant en réalité la voiture en moyenne 0 à 3.5 jours par semaine (sous-ensemble flou A). Tous les conducteurs faisant partie de cette classe auront la possibilité d'obtenir une faible prime.

- La catégorie 2 qui rassemble les conducteurs prenant la voiture en moyenne 2.5 à 5.5 jours par semaine (sous-ensemble flou B). Tous les conducteurs faisant partie de cette classe auront une prime moyenne.
- La catégorie 3 qui unit finalement les conducteurs prenant la voiture en moyenne 4.5 à 7 jours par semaine (sous-ensemble flou C). Tous les conducteurs faisant partie de cette classe auront quant à eux une prime assez élevée.

Un des objectifs est de déterminer la valeur de la variable "Avgdays.week" pour laquelle l'assuré, en choisissant le contrat télématique, obtiendra la même prime pure qu'en choisissant un contrat traditionnel. Cette valeur sera appelée le nombre de base. Si l'assuré estime qu'il utilise en moyenne moins sa voiture que ce nombre de base, celui-ci pourra bénéficier d'une prime plus faible et le contrat télématique sera pour ce dernier très intéressant. À l'inverse, un assuré pensant qu'il utilise sa voiture plus que ce nombre de base ne verra pas d'intérêt à prendre un contrat télématique, car il en sera pénalisé.

Pour déterminer ce nombre de base, la première étape consiste à reconstruire un GLM en ajoutant aux variables classiques cette variable télématique, qui n'est quant à elle pas catégorisée. Les coefficients des variables explicatives sont obtenus dans le tableau 4.1.2. Le coefficient de la variable télématique est bien positif ($\beta'_{16} = 0.105$).

Variable	Coefficient	Variable (suite)	Coefficient
Intercept	$\beta'_0 = -8.784$	Credit.score between 766 and 825	$\beta'_8 = -0.403$
Insured age between 39 and 51 years	$\beta'_1 = 0.115$	Credit.score between 825 and 856	$\beta'_9 = -0.703$
Insured age between 51 and 63 years	$\beta'_2 = 0.079$	Credit score more than 856	$\beta'_{10} = -0.694$
Insured age more than 63 years old	$\beta'_3 = -0.078$	Region Urban	$\beta'_{11} = 0.174$
Car age more than 5 years old	$\beta'_4 = -0.441$	More than 7456.452 annual miles drive	$\beta'_{12} = 0.155$
Car.use Commute	$\beta'_5 = -0.248$	Between 15 and 29 years no claims	$\beta'_{13} = -0.270$
Car.use Farmer	$\beta'_6 = -0.796$	Between 29 and 41 years no claims	$\beta'_{14} = -0.258$
Car.use Private	$\beta'_7 = -0.198$	More than 41 years no claims	$\beta'_{15} = -0.096$
		Average days week	$\beta'_{16} = 0.105$

TABLE 4.1.2 – Coefficients du GLM du contrat télématique

La seconde étape consiste, pour chaque assuré, à évaluer les valeurs prédites par le contrat classique à celles prédites par le contrat télématique, en gardant comme inconnue la variable "Nombre de base" (en tenant compte de l'exposition e , voir théorie à l'équation (1.1.6)). Le nombre de base pour l'assuré qui illustre l'exemple est tel que :

$$-4.399 = \beta'_0 + \beta'_3 + \beta'_4 + \beta'_7 + \beta'_9 + \beta'_{11} + \beta'_{15} + (\beta'_{16} \times \text{Nombre de base}) + \log(e)$$

$$\rightarrow \text{Nombre de base} = \frac{(-4.399 - \beta'_0 - \beta'_3 - \beta'_4 - \beta'_7 - \beta'_9 - \beta'_{11} - \beta'_{15} - \log(e))}{\beta'_{16}} = 4.918$$

Par conséquent, afin d'obtenir une prime pure du contrat télématique égale à la prime pure du contrat classique, l'assuré doit utiliser sa voiture en moyenne 4.918 jours par semaine. Dans ce cas, si l'assuré estime qu'il utilise sa voiture moins de fois que ce nombre

de base, le contrat télématique pourrait potentiellement lui offrir une prime plus faible.

Selon sa connaissance subjective, l'assuré décide de choisir une des trois catégories. Par exemple, si ce dernier estime qu'il n'utilise que très peu sa voiture et qu'il se trouve dans la première catégorie, alors l'assureur estime qu'il y a 80% de chance qu'il l'utilise vraiment un nombre de jours moyen allant entre 0 et 3. L'assureur répartit les 20% restant de manière équitable dans les deux autres catégories. Selon l'assureur, l'assuré appartient donc à 80% au sous-ensemble flou A et à 10% aux sous-ensembles flous B et C. Ceci permet à l'assureur de tenir compte de la subjectivité de l'assuré et de rester prudent. Pour chaque nombre flou, le milieu du segment représentant l'alpha-coupe (voir sous-section 2.3.2) est choisi comme nombre moyen de jours où le véhicule est utilisé. Les sous-ensembles A, B et C représentant la variable *Avgdays.week* sont illustrés à la figure 4.1.1.

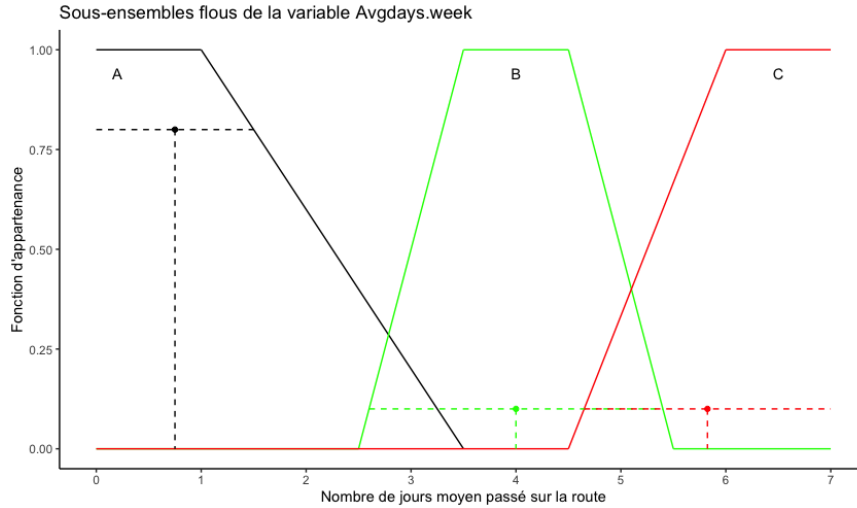


FIGURE 4.1.1 – Estimation par l'assureur du nombre de jours d'usage réel du véhicule d'un assuré (exemple)

La valeur finale du nombre moyen d'utilisations du véhicule par semaine estimée par l'assureur pour cet individu, notée $Avgdays_{est}$, est calculée comme suit :

$$\begin{aligned} Avgdays_{est} &= 80\% \times \frac{(\bar{a} - \underline{a})}{2} + 10\% \times (\underline{b} + \frac{(\bar{b} - \underline{b})}{2}) + 10\% \times (\underline{c} + \frac{(\bar{c} - \underline{c})}{2}) \\ &= 80\% \times 0.75 + 10\% \times 4 + 10\% \times 5.825 = 1.58 \end{aligned}$$

La définition d'un nombre flou trapézoïdal permet de trouver les valeurs des bornes de l'alpha-coupe de chaque sous-ensemble, à savoir \bar{a} , \underline{a} , \bar{b} , \underline{b} ,... (voir sous-section 2.4 équation 2.4.3). Par exemple, pour trouver \bar{a} , il faut chercher la valeur inconnue de x qui résout l'équation égalisant 80% au côté droit du sous-ensemble A :

$$0.8 = \frac{x - 3.5}{1 - 3.5} \rightarrow x = 1.5$$

Pour le sous-ensemble A, l'équation 4.1.2 de $V_P^\lambda(A)$ est appliquée pour $\lambda = 0.5$ et une probabilité constante sur α posée à 80%. Ce choix est arbitraire et une loi centrée autour de 80% aurait aussi pu être choisie. Comme l'assuré a décidé d'appartenir à la catégorie 1, l'assureur estime qu'il utilise sa voiture en moyenne 1.58 jour par semaine. Ce nombre étant inférieur au nombre de base, l'assuré mérite une réduction de prime et aurait donc tout intérêt à choisir le contrat télématique. En effet, la fréquence de sinistres égale à 0.00865 est bien inférieure à 0.01229. Le tableau 4.1.3 synthétise tous les résultats. Si l'assuré décide de se classer dans la catégorie 3, il n'a pas intérêt à prendre de contrat télématique, car il devra payer une prime plus élevée. À l'inverse, si l'assuré avait plutôt choisi d'appartenir à la catégorie 2, il aurait été intéressant pour lui de prendre un contrat télématique pour réduire sa prime.

Choix de l'assuré	$Avgdays_{est}$	Prime pure contrat traditionnel	Prime pure contrat télématique	Choix du contrat
Catégorie 1	1.58	0.01229	0.00865	télématique
Catégorie 2	3.94	0.01229	0.01109	télématique
Catégorie 3	5.64	0.01229	0.01326	traditionnel

TABLE 4.1.3 – Résultats des choix de l'assuré selon différents scénarios

4.2 L'espérance floue selon Dubois et Prade

4.2.1 Théorie

Une définition pertinente de la valeur moyenne d'un nombre flou, cohérente avec la définition classique de la théorie des probabilités, a été proposée par Dubois et Prade (1987). Si A est un intervalle flou semi-continu supérieur et $\mathcal{P}(A)$ représente l'ensemble des mesures de probabilité sur le support $S(A)$ (voir sous-section 2.3.3), alors la valeur moyenne de A, notée $E(A)$, est un ensemble de nombres défini par l'équation (4.2.1).

$$E(A) = \{E(P) | P \in \mathcal{P}(A)\} = [E_*(A), E^*(A)] \quad (4.2.1)$$

Où $E(P)$ est l'espérance classique défini par l'équation (4.2.2).

$$E(P) = \int_{-\infty}^{+\infty} x dF(x) \quad (4.2.2)$$

F étant la fonction de distribution de P. De plus, $E_*(A) = \inf(E(A))$ et $E^*(A) = \sup(E(A))$, représentent respectivement la valeur moyenne inférieure et supérieure de A. Comme pour tout $P \in \mathcal{P}(A)$, $F_* \leq F \leq F^*$, l'équation (4.2.3) définit $E_*(A)$ et $E^*(A)$.

$$E_*(A) = \int_{-\infty}^{+\infty} x dF^*(x); E^*(A) = \int_{-\infty}^{+\infty} x dF_*(x) \quad (4.2.3)$$

Où $F^*(x)$ et $F_*(x)$ représentent respectivement la fonction de distribution supérieure et inférieure. Selon Dubois et Prade (1987), celles-ci peuvent être définies par les équations présentées en (4.2.4).

$$F^*(x) = \sup\{A(r) | r \leq x\} ; \quad F_*(x) = \inf\{1 - A(r) | r > x\} \quad (4.2.4)$$

Par conséquent, la valeur moyenne d'un ensemble flou A est un intervalle fermé limité par les valeurs attendues calculées à partir de ses fonctions de distribution. Le lemme 1 permet d'interpréter géométriquement ces propriétés.

Lemme 1. *Soit F la fonction de distribution de P. Alors, dès que :*

$$\lim_{x \rightarrow -\infty} x^k F(x) = \lim_{x \rightarrow +\infty} x^k (1 - F(x)) = 0$$

Pour tout $k > 1$, il en suit que :

$$E[P] = \int_0^{+\infty} (1 - F(x)) dx - \int_{-\infty}^0 F(x) dx$$

Soit un ensemble flou A tel que le noyau soit un intervalle borné ($C(A) = [a^-, a^+]$, voir sous-section 2.3.4) et tel que F^* et F_* satisfont aux hypothèses du lemme 1, alors $E_*(A)$ et $E^*(A)$ sont définies par les équations présentes en (4.2.5).

$$\begin{aligned} E_*(A) &= \int_{-\infty}^{+\infty} x dF^*(x) = \int_0^{+\infty} (1 - F^*(x)) dx - \int_{-\infty}^0 F^*(x) dx \\ &= a^- - \int_{-\infty}^{a^-} A(x) dx \end{aligned} \quad (4.2.5)$$

$$\begin{aligned} E^*(A) &= \int_{-\infty}^{+\infty} x dF_*(x) = \int_0^{+\infty} (1 - F_*(x)) dx - \int_{-\infty}^0 F_*(x) dx \\ &= a^+ + \int_{a^+}^{+\infty} A(x) dx \end{aligned}$$

4.2.2 Lien avec la théorie présentée par Gonzalez et de Campos via la mesure de Lebesgue

Il est possible de faire un lien entre la définition de l'espérance floue selon Dubois et Prade (sous-section 4.2.1) et la classification vue par de Campos et Gonzalez (sous-section 4.1.1) en utilisant la mesure de Lebesgue. La mesure de Lebesgue sur $[0, 1]$, notée P_L , est telle que $P_L((a, b]) = b - a$, $(a, b] \in \mathcal{P}([0, 1])$ (de Campos Ibáñez et González Muñoz, 1989).

Soit R^* l'ensemble des nombres flous normalisés A qui ont une fonction caractéristique $A(x)$ continue et strictement croissante avant d'atteindre les valeurs unitaires (partie $r_A(x)$) et qui décroît strictement après celles-ci (partie $s_A(x)$), telle que dans l'équation (4.2.6).

$$A(x) = \begin{cases} r_A(x) & \text{si } x \in [a, b] \\ 1 & \text{si } x \in [b, c] \\ s_A(x) & \text{si } x \in [c, d] \\ 0 & \text{sinon} \end{cases} \quad (4.2.6)$$

González (1990) prouve que $\forall A \in R^* : E(A) = \{V_{P_L}^\lambda(A) | \lambda \in [0, 1]\}$. Cette démonstration ne fonctionne que si le nombre flou détient une fonction caractéristique semblable à l'équation (4.2.6).

Démonstration. Si $A \in R^*$ et $A^\alpha = [a, \bar{a}]$, alors :

$$E_*(A) = \int_{-\infty}^{+\infty} x dF^*(x) \text{ et } E^*(A) = \int_{-\infty}^{+\infty} x dF_*(x)$$

Avec :

$$F^*(x) = \sup\{A(r) | r \leq x\} = \begin{cases} r_A(x) & \text{si } x < b \\ 1 & \text{si } x \geq b \end{cases}$$

$$F_*(x) = \inf\{1 - A(r) | r > x\} = \begin{cases} 0 & \text{si } x < c \\ 1 - s_A(x) & \text{si } x \geq c \end{cases}$$

Dès lors sont obtenues les équations :

$$E_*(A) = \int_{-\infty}^{+\infty} x dF^*(x) = \int_a^b x dr_A(x) = \int_0^1 r_A^{-1}(x) dx = \int_0^1 a d\alpha = V_{P_L}^0(A)$$

$$E^*(A) = \int_{-\infty}^{+\infty} x dF_*(x) = - \int_c^d x ds_A(x) = \int_0^1 s_A^{-1}(x) dx = \int_0^1 \bar{a} d\alpha = V_{P_L}^1(A)$$

Par la proposition 1 présente à la sous-section 4.1.1, $V_{P_L}^\lambda(A)$ est tel que :

$$\forall \lambda \in [0, 1] : V_{P_L}^\lambda(A) = \lambda V_{P_L}^1(A) + (1 - \lambda)V_{P_L}^0(A)$$

□

Si A est un nombre flou de forme triangulaire tel que $A = Tr[a, b, c]$ (voir sous-section 2.4.1), l'équation (2.4.2) permet de déterminer son espérance floue qui est telle que :

$$V_{P_L}^\lambda(A) = \lambda \int_0^1 \overline{Tr[a, b, c]} d\alpha + (1 - \lambda) \int_0^1 \underline{Tr[a, b, c]} d\alpha$$

$$= \lambda \int_0^1 b + (1 - \alpha)(c - b) d\alpha + (1 - \lambda) \int_0^1 b - (1 - \alpha)(b - a) d\alpha^1$$

1. $V_{P_L}^\lambda(A) = \lambda(\frac{b}{2} + \frac{c}{2}) + (1 - \lambda)(\frac{b}{2} + \frac{a}{2})$

Si A est un nombre flou de forme trapézoïdale, le même raisonnement est applicable, en utilisant l'équation (2.4.4) au lieu de l'équation(2.4.2) ².

4.2.3 Application : approche floue des intervalles de confiance

L'exemple précédent aidait le choix de l'assuré concernant le type de contrat d'assurance. Le concept de l'espérance selon Dubois et Prade peut quant à lui s'appliquer aux décisions de l'assureur. Une application en trois étapes permet de déterminer s'il est intéressant pour ce dernier de proposer un contrat télématique à son assuré. Il est important pour l'assureur d'identifier les bons clients afin de leur proposer le meilleur contrat.

Premièrement, l'assureur détermine un intervalle qui correspond aux bons conducteurs en construisant un sous-ensemble flou sur base d'une variable télématique. Il définit ensuite les bornes de l'espérance, E_* et E^* telles que définies à l'équation 4.2.3, du sous-ensemble destiné aux bons clients. Si l'assuré détient une variable télématique dont la valeur appartient à cet intervalle, alors il est dans l'intérêt de l'assureur de lui offrir un contrat télématique. L'assureur lui fait un cadeau afin qu'il reste dans son portefeuille. Deuxièmement, le même procédé est applicable au sous-ensemble flou correspondant aux mauvais conducteurs. Si l'assuré détient une variable télématique dont la valeur appartient à cet intervalle, alors il n'est pas dans l'intérêt de l'assureur de lui proposer un tel contrat. Enfin, en ce qui concerne les conducteurs dont la valeur de leur variable télématique n'appartient à aucun des deux intervalles, l'assureur est libre de faire ce que bon lui semble. En théorie, il n'est pas obligé de donner directement un avantage télématique et peut par exemple offrir le produit de base pendant 1 an avant de constater les avantages qu'il peut offrir.

Comme exemple simplifié, la variable "Avgdays.week" de la base de données (voir annexe A.1) est utilisée afin de déterminer si un assuré est un bon conducteur. Cette variable télématique est la même que celle utilisée pour l'exemple précédent (voir sous-section 4.1.2) et décrit le nombre de jours moyens passés sur la route. Le sous-ensemble A construit à la figure 4.1.1 est repris pour cet exemple. Cet ensemble regroupe les assurés qui ont un bon profil de risque en utilisant leur voiture en moyenne 1 à 3.5 jours par semaine. La première étape consiste à déterminer la fonction caractéristique $A(r)$ de ce sous-ensemble A :

$$A(r) = \begin{cases} 0 & \text{si } r < 0 \text{ ou } r \geq 3.5 \\ 1 & \text{si } 0 \leq r < 1 \\ \frac{r-3.5}{1-3.5} & \text{si } 1 \leq r < 3.5 \end{cases}$$

Grâce à la fonction d'appartenance $A(r)$ et à son complément, il y a lieu pour tout x , de calculer $F^*(x)$ et $F_*(x)$ selon les formules présentent en (4.2.4). Concernant $F^*(x)$, il faut prendre pour tout x le degré d'appartenance maximal contenu dans l'ensemble des $r \leq x$. Pour $F_*(x)$, il faut choisir pour tout x le complément du degré d'appartenance minimum contenu dans l'ensemble $r > x$. Le tableau 4.2.1 résume l'information nécessaire

2. Dans ce cas, $V_{PL}^\lambda(A) = \lambda \int_0^1 c + (1 - \alpha)(d - c) d\alpha + (1 - \lambda) \int_0^1 b - (1 - \alpha)(b - a) d\alpha$

à la construction des fonctions de distribution. La figure 4.2.1 illustre la fonction d'appartenance $A(r)$ ainsi que son complément et les fonctions de répartition F^* et F_* .

x	$r \leq x$	$\sup\{A(r) r \leq x\}$	$r > x$	$\inf\{1 - A(r) r > x\}$
0	0	1	$]0,7]$	0
1	$[0,1]$	1	$]1,7]$	0
3	$[0,3]$	1	$]3,7]$	0.8
5	$[0,5]$	1	$]5,7]$	1
7	$[0,7]$	1	$]7,7]$	1

TABLE 4.2.1 – Quelques valeurs utiles à la construction des fonctions F^* et F_* pour le sous-ensemble A

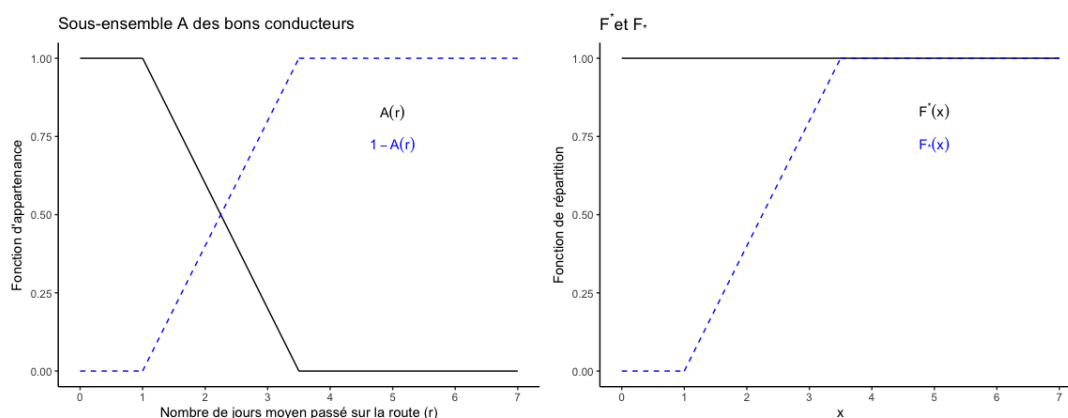


FIGURE 4.2.1 – Sous-ensemble A représentant les bons conducteurs et les fonctions de distribution supérieure (F^*) et inférieure (F_*)

Comme le noyau de A est $C(A) = [0, 1]$ (voir sous-section 2.3.4), $a^- = 0$ et $a^+ = 1$. Par conséquent, en utilisant les équations présentées en (4.2.5), les bornes de l'espérance du sous-ensemble flou A sont :

$$E_*(A) = 0 \text{ et } E^*(A) = 1 + \int_1^{3.5} \frac{x - 3.5}{-2.5} dx = 2.25$$

L'intervalle flou de confiance $[E_*(A), E^*(A)]$ est donc compris entre 0 et 2.25. Dès lors, si un assuré détient une variable "Avgdays.week" dont la valeur est comprise entre cet intervalle, il sera dans l'intérêt de l'assureur de lui proposer un contrat télématique.

Le même raisonnement doit être réalisé pour le sous-ensemble C (voir figure 4.1.1) qui correspond aux conducteurs plus à risques. Le développement se trouve en annexe C. Comme résultat, l'intervalle de confiance $[E_*(C), E^*(C)]$ est compris entre 5.25 et 7. Par

conséquent, si un assuré détient une variable "Avgdays.week" dont la valeur est supérieure ou égale à 5.25, il sera dans l'intérêt de l'assureur de ne pas lui proposer un contrat télématique.

Ces deux exemples illustrent l'effet d'une seule variable, à savoir le nombre moyen de jours passés sur la route, mais il est bien entendu possible de cumuler l'effet de nombreuses variables télématiques au sein d'un même contrat. Pour ce faire, il est par exemple possible de créer une variable déterminant un score de conduite pour chaque assuré sur base de plusieurs variables télématiques. Le chapitre 5 sera consacré à la construction d'un score flou de conduite en intégrant plus d'une variable télématique. Ce chapitre utilisera la théorie de la logique floue présente au chapitre 3.

4.3 L'espérance floue selon Kandel

Une autre manière³ de calculer l'espérance floue fut élaborée par [Kandel et Byatt \(1978\)](#). Dans la théorie de probabilité classique, un système de probabilité est un triplet (Ω, S, P) où Ω est un ensemble arbitraire regroupant tous les résultats possibles, S est une collection d'événements A et P est une fonction à valeur réelle définie pour chaque $A \in S$. Ce système répond à trois axiomes de probabilités, à savoir :

1. La probabilité P d'un événement A est représentée par un nombre réel compris entre 0 et 1 : $0 \leq P(A) \leq 1$
2. La probabilité d'obtenir un quelconque résultat de l'univers est égale à 1 : $P(\Omega) = 1$
3. Si A_1, A_2, \dots est une séquence d'événements deux à deux disjoints dans S , alors : $P(A_1 \cup A_2 \cup \dots) = \sum_{i=1}^{+\infty} P(A_i)$

Selon Kandel, cette troisième propriété est critiquable, car elle implique une certaine exhaustivité qui restreint l'emplacement de chaque point d'échantillonnage dans un ensemble bien défini. Bien entendu, pour les espaces finis, la probabilité de tout événement A , $P(A)$, respecte l'axiome numéro 3. Cependant, ce mémoire illustre clairement le fait qu'un événement peut être flou s'il appartient à un ensemble flou. De plus, même si un événement est bien défini, sa probabilité peut quant à elle être mal définie suite à des évaluations peu précises, par exemple. Dès lors, l'objectif est de trouver une mesure alternative qui supprime l'additivité et qui s'applique aux données floues.

La notion de mesure floue est définie par [Kandel et Byatt \(1978\)](#) de la manière suivante. Soit B une tribu borélienne (σ -algèbre) de sous-ensembles présents dans Ω . Un ensemble de fonctions défini sur B , noté $\mu(\cdot)$, est appelé mesure floue s'il détient les propriétés suivantes :

1. $\mu(\Phi) = 0$
2. $\mu(\Omega) = 1$

3. Pour rappel, la section 4.3 et suivantes sont là afin de montrer la grande variété des outils flous et une étude plus approfondie n'a pas été effectuée.

3. Si $A, B \in B$ avec $A \subset B$, alors $\mu(A) \leq \mu(B)$
4. Si $\{x_j | 1 \leq j < \infty\}$ est une suite monotone, alors $\lim_{j \rightarrow +\infty} [\mu(x_j)] = \mu[\lim_{j \rightarrow +\infty} (x_j)]$. Clairement $\Omega, \Phi \in B$ et si $x_j \in B$ et $\{x_j | 1 \leq j < \infty\}$ est une suite monotone, alors $\lim_{j \rightarrow +\infty} (x_j) \in B$

Soit $A : \Omega \rightarrow [0, 1]$ une fonction caractéristique où $A^\alpha = \{x | A(x) \geq \alpha\}$ représente l'alpha-coupe (voir sous-section 2.3 du chapitre 2). La fonction A est une fonction B -mesurable si $A^\alpha \in B \forall \alpha \in [0, 1]$. Si A est une fonction mesurable de B telle que $A \in [0, 1]$, alors la valeur attendue floue de A sur un ensemble A' par rapport à la mesure floue $\mu(\cdot)$, notée $FEV(A)$, est définie par l'équation (4.3.1).

$$FEV(A) = \sup_{\alpha \in [0,1]} \{ \min[\alpha, \mu(A^\alpha)] \}, \quad \text{Où } A^\alpha = \{x | A(x) \geq \alpha\} \subseteq A' \quad (4.3.1)$$

La fonction $f_A(\alpha) = \mu(x | A(x) \geq \alpha)$ dépend du niveau α . Le calcul de l'espérance floue dans ce cas consiste donc à trouver l'intersection de la courbe telle que $\alpha = f_A(\alpha) = H$, de sorte que $FEV(A) = H \in [0, 1]$.

Un exemple de cette théorie se trouve dans un article écrit par [Schneider et Kandel \(1988\)](#). Ce concept utilise le principe d'extension de Zadeh décrit à la sous-section 2.5.1 du chapitre 2, ce qui prouve encore une fois que c'est un principal fondamental. Dans un article, [Kandel et Byatt \(1978\)](#) a également défini l'extension de cette définition quand $A \in [a, b]$, $0 \leq a < b < \infty$.

4.4 L'espérance floue dans un espace de crédibilité

Pour cette sous-section, il est nécessaire de présenter deux nouvelles mesures floues, à savoir la mesure de possibilité et la mesure de nécessité ([Baoding Liu et Yian-Kui Liu, 2002](#)). Si r est un nombre réel et A est une variable aléatoire floue normalisée, le degré de possibilité que $\{A \geq r\}$ se définit par l'équation (4.4.1). Le degré de nécessité que $\{A \geq r\}$ peut lui se définir par l'équation (4.4.2).

$$Pos\{A \geq r\} = \sup_{u \geq r} \{A(u)\} \quad (4.4.1)$$

$$Nec\{A \geq r\} = 1 - Pos\{A < r\} = 1 - \sup_{u < r} \{A(u)\} \quad (4.4.2)$$

Sur base de ces deux mesures, l'indice de crédibilité peut se définir par l'équation (4.4.3).

$$Cr(A) = \frac{1}{2}(Pos\{A\} + Nec\{A\}) \quad (4.4.3)$$

Pos et Nec représentent ensemble une paire de mesure floue, car il est possible de définir l'une des deux en utilisant l'autre, tel que $Pos\{A\} = 1 - Nec\{A^C\}$ où A^C représente le

complément de A . Dès lors, si A est une variable aléatoire normalisée floue, les valeurs attendues supérieures et inférieures de A , notées respectivement $\bar{E}[A]$ et $\underline{E}[A]$, sont définies par l'équation (4.4.4). La définition de l'intégrale de Choquet (1954) motive ces définitions.

$$\begin{aligned}\bar{E}[A] &= \int_0^\infty Pos\{A \geq r\} dr - \int_{-\infty}^0 Nec\{A \leq r\} dr \\ \underline{E}[A] &= \int_0^\infty Nec\{A \geq r\} dr - \int_{-\infty}^0 Pos\{A \leq r\} dr\end{aligned}\quad (4.4.4)$$

L'espérance floue de A , notée $E[A]$, est définie par l'équation (4.4.5).

$$E[A] = \int_0^\infty Cr\{A \geq r\} dr - \int_{-\infty}^0 Cr\{A \leq r\} dr \quad (4.4.5)$$

Cette espérance existe à condition qu'au moins une des deux intégrales soit finie.

4.5 L'espérance floue selon Puri et Ralescu

L'intégrale de Aumann (1965) a été généralisée par Puri et Ralescu (1986) afin de définir l'espérance d'une variable aléatoire floue. Selon eux, soit un espace de probabilité (Ω, \mathcal{A}, P) , χ est une variable aléatoire floue si elle est D_W^ϕ -Borel mesurable. Puri and Ralescu définissent la valeur attendue floue comme une mesure de la tendance centrale d'une variable aléatoire floue de la manière suivante. Si χ est une variable aléatoire floue telle que la variable aléatoire $\max\{|\inf\chi^0|, |\sup\chi^0|\}$ est intégrable, alors la valeur attendue floue de χ correspond à $E(\chi)$ (équation (4.5.1)).

$$(E(\chi))^\alpha = [E(\inf\chi^\alpha), E(\sup\chi^\alpha)] \quad \forall \alpha \in [0, 1] \quad (4.5.1)$$

Cette espérance floue de χ est définie comme étant l'unique élément tel que $\forall \alpha \in [0, 1]$:

$$\begin{aligned}(E(\chi))^\alpha &= \text{l'intégrale d'Aumann (1965) de l'ensemble aléatoire } \chi^\alpha \\ &= \left\{ \int_\Omega X(\omega) dP(\omega) \mid X : \Omega \rightarrow R, X \in L^1(\Omega, \mathcal{A}, P), X \in \chi^\alpha \right\}\end{aligned}$$

$L^1(\Omega, \mathcal{A}, P)$ dénote l'espace des fonctions P -intégrables $X : \Omega \rightarrow R$. Par ailleurs, la variance d'une variable aléatoire floue peut être définie grâce à la notion de distance entre deux nombres flous D_W^ϕ définie à la sous-section 2.5.3 (Lubiano et al., 2000). Soit χ une variable aléatoire floue telle que $[\max\{|\inf\chi^0|, |\sup\chi^0|\}]^2$ est intégrable, alors, la variance de χ est donnée par l'équation (4.5.2).

$$Var(\chi) = E([D_W^\phi(\chi, E(\chi))]^2) = \int_\Omega [D_W^\phi(\chi(\omega), E(\chi))]^2 dP(\omega) \quad (4.5.2)$$

Chapitre 5

Construction d'un produit télématique

5.1 Base de données

Ce chapitre s'appuie sur une base de données simulée qui est présentée en annexes [A.1](#). Les résultats présentés ci-dessous sont donc donnés à titre d'exemple afin d'illustrer et d'appliquer la théorie présente dans ce mémoire. Il serait dès lors intéressant de réaliser ces analyses sur une base de données réelle afin de tirer des résultats fiables et utiles pour le domaine des assurances. L'utilisation de cette base de données a pour principal objectif d'analyser la fréquence de sinistres. Au sein d'un portefeuille automobile d'un assureur, il y a généralement, au cours d'une année de police, une grande proportion de conducteurs qui n'ont causé aucun sinistre, une proportion moindre avec exactement un sinistre et une proportion bien moindre avec deux sinistres ou plus. Cet effet peut s'amplifier lors d'une étude sur la télématique en secteur automobile, car ce sont généralement les personnes qui se sentent les plus prudentes sur la route qui se dirigent vers ce genre d'assurance. Dans la partie de la base de données dédiée à l'entraînement du modèle, il y a 95.73% d'observations sans sinistre, 4.06% avec exactement 1 sinistre, 0.2% avec 2 sinistres et 0.01% avec 3 sinistres. Dans l'analyse, il faut donc être conscient que les observations sont assez déséquilibrées.

5.2 Construction de plusieurs sous-scores flous de conduite à l'aide de la logique floue

Cette sous-section est dédiée à la construction, pour chaque assuré, d'un score de conduite compris entre 0 et 100 qui dépend de toutes les variables télématiques présentes dans la base de données (voir annexe [A.1](#)). Au plus le score se rapproche de 100, au plus l'assuré sera estimé comme un conducteur à faible risque. Pour ce faire, la logique floue présentée au chapitre [3](#) est utilisée. Cependant, comme des règles floues (voir sous-section [3.1.3](#)) doivent-être élaborées afin de construire ces sous-scores, il est difficilement concevable de créer un score flou unique comprenant les 39 variables télématiques de la base de

données. En effet, le nombre de règles floues qui sous-tendrait ce système serait beaucoup trop compliqué à mettre en place. Par conséquent, le choix s'est tourné vers la construction de plusieurs sous-scores dérivant chacun d'un plus petit nombre de variables. Ces derniers seront ensuite utilisés afin de modifier la tarification des primes. Plusieurs étapes sont nécessaires, à savoir une analyse de corrélation et une analyse en composantes principales (ACP) (voir sous-section 5.2.1), l'étude du lien entre les regroupements effectués par l'ACP et la sinistralité (voir sous-section 5.2.2) et enfin la création d'un système d'inférence floue pour chacun des sous-scores déterminés (voir sous-section 5.2.3).

5.2.1 Corrélation et analyse en composantes principales (ACP)

Tout d'abord, une analyse de corrélation est effectuée afin de supprimer les variables très corrélées entre elles. Le coefficient de Pearson, mesurant la corrélation linéaire entre deux variables, est utilisé. Les variables qui présentent une corrélation supérieure à 60% en valeur absolue sont retirées du modèle et seules 11 variables sont retenues (voir tableau D.1.1 en annexe D.1).

L'étape suivante consiste à déterminer plusieurs catégories de variables qui définiront chacune un sous-score de conduite. Pour ce faire, une analyse en composantes principales (ACP) est réalisée. Cette analyse étudie les relations qui existent entre des variables quantitatives et permet de résumer l'information en un nombre réduit de nouvelles variables appelées composantes principales (Lebart et al., 2006). Chaque groupe de variables déterminé servira au calcul d'un sous-score de conduite concernant un domaine bien précis (par exemple la vitesse). Le nombre de composantes principales est inférieur ou égal au nombre de variables explicatives (11). Bien entendu, l'ACP n'est pas la seule façon de regrouper des variables entre elles, mais c'est cette dernière qui a été choisie pour ce mémoire.

Dans R, la fonction *PCA()* (package FactoMineR) permet d'effectuer cette analyse. En annexe D.1, le cercle de corrélation pour les dimensions 1 et 2 ainsi que la qualité de représentation des variables avec 5 dimensions sont présents à la figure D.1.2. Le cercle de corrélation détermine les variables qui sont fortement corrélées entre elles, tandis que la qualité de représentation met en évidence les variables les plus contributives pour chaque dimension. Le cercle de corrélation révèle que les variables "Pct.drive.2hrs" et "Pct.drive.4hrs" sont négativement corrélées à la dimension 1, tandis que les variables "Accel.06miles" et "Brake.06miles" en sont le plus positivement corrélées. Intuitivement, il est facile de comprendre qu'au plus la voiture est utilisée pour de longues durées, au moins l'assuré accélère et freine brusquement. Deux profils peuvent ressortir, à savoir celui qui roule plutôt agressivement ou celui qui, étant habitué aux longs trajets, roule de manière plus calme. En ce qui concerne la dimension 2, ces quatre variables en sont positivement corrélées.

À première vue, les regroupements de variables réalisés par l'ACP semblent évidents. Par exemple, les variables "Accel.06miles" et "Brake.06miles" appartiennent à la catégorie vitesse, tandis que les variables "Pct.drive.2hrs" et "Pct.drive.4hrs" représentent le pour-

centage de véhicules conduits pendant une longue durée (2 et 4h). La conduite en heures de pointe peut être résumée par les variables "Pct.drive.rush.am" et "Pct.drive.rush.pm". La dimension 4, expliquée principalement par les variables "Pct.drive.wkday", "annual.pct.driven" et "avgdays week" peut exprimer le temps passé sur la route. Finalement, l'information liée à la prise des virages peut être résumée par la dimension 5, par les variables "Left.turn.intensity08" and "Right.turn.intensity08". Ces cinq groupes de variables sont des candidats potentiels à la création de cinq sous-scores de conduite. Avant cela, il est toutefois nécessaire d'évaluer s'il existe un lien entre ces derniers et la sinistralité (voir sous-section 5.2.2).

5.2.2 Lien entre les regroupements effectués par l'ACP et la sinistralité

En proposant un produit télématique, l'assureur fait face à un dilemme. D'un côté, son produit doit attirer l'assuré en permettant une variation de prime en fonction du comportement. De l'autre, l'assureur ne veut pas faire de cadeaux et souhaite que cela ait une conséquence sur la sinistralité. Le changement de comportement de l'assuré doit permettre une diminution du risque sur les routes. Cette sous-section consiste à trouver du sens aux regroupements de variables établis à la sous-section 5.2.1 en analysant la relation entre ces groupes de variables et la sinistralité. Pour ce faire, un GLM est réalisé (voir théorie à la sous-section 1.1.1) avec comme variable à expliquer le nombre de sinistres et comme variables explicatives les variables classiques ainsi que les cinq premières dimensions de l'ACP réalisée (modèle 1). Ce GLM est ensuite comparé à un GLM comprenant les variables classiques ainsi que les 11 variables télématiques non corrélées (modèle 2). Les résultats sont résumés dans le tableau 5.2.1.

Modèle	AIC	Déviante	Variabes télématiques/dimensions non significatives
1	22 200	16 700	/
2	20 670	15 160	Right.turn.intensity08

TABLE 5.2.1 – Résultats des différents GLM comprenant les dimensions de l'ACP

Résumer fait perdre de l'information. Dès lors, il n'est pas surprenant de voir que le modèle 2 donne de meilleurs résultats que le modèle 1, sachant que les 5 dimensions n'expliquent que 61.8% de la variance totale (voir scree plot en annexe D.1 figure D.1.1). Cependant, il n'est pas gérable de construire un score flou sur base d'autant de variables et regrouper l'information de façon logique est nécessaire. De plus, pour l'aspect marketing, construire un score de conduite sur base de 11 variables restera assez vague pour la majorité des assurés. La combinaison de plusieurs sous-scores avec un nom parlant semble être le meilleur choix. Selon le premier modèle, les cinq dimensions sont significatives. La variable "Right.turn.intensity08" n'est pas significative selon le modèle 2. L'aspect statistique qui considère le modèle 2 comme étant plus optimal s'oppose à l'aspect marketing qui considère le regroupement de plusieurs variables comme étant plus compréhensible. Comme c'est un

produit risko-marketing, le choix final se tourne vers la création de cinq sous-scores de conduite présents dans le tableau 5.2.2.

1 : Vitesse	2 : Heures de conduite	3 : Prise des virages
Accel.06miles	Pct.drive.2hrs	Left.turn.intensity08
Brake.06miles	Pct.drive.4hrs	Right.turn.intensity08
3 : Conduite en heures de pointe	4 : Temps passé sur la route	
Pct.drive.rush.am	Pct.drive.wkday	
Pct.drive.rush.pm	Annual.pct.driven	
	Avgdays week	

TABLE 5.2.2 – Présentation des cinq sous-scores de conduite liés à la sinistralité

Même si ce choix n'est pas idéal statistiquement, il ne faut pas oublier que les données sont simulées. De plus, avec le temps, l'utilisation de ces sous-scores de conduite peut faire varier les comportements routiers et ces groupes de variables peuvent prendre de plus en plus d'importance. Si l'objectif est de construire un produit télématique transparent envers le client, ces scores doivent pouvoir lui être expliqués et celui-ci doit les comprendre.

5.2.3 Création d'un système d'inférence floue pour chacun des sous-scores déterminés

La dernière étape consiste à créer un système d'inférence floue (voir section 3.1 du chapitre 3) pour chacun des sous-scores de conduite. Pour ce faire, le package FuzzyR (voir annexe B) est utilisé dans R. Comme les étapes nécessaires à la création d'un système d'inférence floue sont assez similaires pour chaque sous-score, les fonctions d'appartenances et les règles floues ont été placées en annexe D.2. Pour le sous-score "temps passé sur la route", l'élaboration des règles floues avec trois variables était compliquée à mettre en place sans l'avis d'experts. Dès lors, seules les variables "Annual pct.driven" et "Avgdays.week" ont été prises en compte, car ce sont les plus significatives.

Des nombres flous de formes trapézoïdales ont été utilisés afin de construire les fonctions d'appartenance des ensembles flous d'entrée et de sortie. La construction des sous-ensembles flous des variables d'entrée tient compte des analyses exploratoires. Par exemple, comme le nombre d'accidents augmente avec le nombre d'accélération (variable d'entrée "Accel.06miles" du sous-score vitesse), les nombres flous ont été construits sur base des quantiles (voir tableau D.2.1 et figure D.2.1 de l'annexe D.2.1). En ce qui concerne les règles floues, l'opérateur flou choisi parmi tous les antécédents est le "ET". C'est le cas de figure où les variables doivent être observées simultanément afin de sélectionner une valeur de sortie. Si une règle possède deux antécédents au sein d'une catégorie, le résultat de la conséquence va toujours compenser les deux effets.

Le tableau 5.2.3 résume l'ensemble des résultats relatifs aux sous-scores de conduite pour l'ensemble des assurés du portefeuille d'apprentissage.

Sous-score	Min	Max	Moyenne
Vitesse	10.56	89.44	48.78
Heures de conduite	10.56	89.44	75.81
Conduite en heures de pointe	12.32	89.44	53.06
Temps passé sur la route	10.56	89.44	43.08
Prise des virages	13.02	89.44	57.23

TABLE 5.2.3 – Résultats des sous-scores par catégorie

5.3 Conception du produit télématique : le système bonus-malus

Cette étape consiste à imaginer un produit télématique sur base des sous-scores flous de conduite construits à la sous-section 5.2. Ce produit récompensera les clients qui ont un bon score en leur accordant une réduction de prime et au contraire pénalisera les assurés qui en détiennent un mauvais en augmentant leur prime. Chaque assuré recevra donc pour chaque sous-score une augmentation ou une diminution de prime. En combinant toutes ces variations, le changement de prime final est obtenu. Les questions sont maintenant de savoir à partir de quels scores ces récompenses/sanctions seront déclenchées et de quels montants seront ces dernières. De plus, il est nécessaire de déterminer comment ces variations de prime, résultant de chaque sous-score, seront combinées afin d'obtenir l'augmentation ou la diminution finale de la prime de chaque assuré. La sous-section 5.3.1 est dédiée à la création du système, tandis que la sous-section 5.3.2 illustre ce produit sur base de deux sous-scores. Finalement, la sous-section 5.3.3 tient compte de l'équilibre du portefeuille.

5.3.1 Mise en place du système

Afin de déterminer à partir de quels scores l'assuré reçoit une augmentation ou une diminution de prime, deux approches sont abordées. La première utilise les intersections des sous-ensembles flous représentant le niveau de risque d'un conducteur. Sur la figure 5.3.1 de gauche, le score 25 représente l'intersection entre le sous-ensemble C_1 dédié aux scores qui qualifient un conducteur agressif et le sous-ensemble C_2 dédié aux conducteurs plus modérés. Le score 75 représente quant à lui l'intersection entre le sous-ensemble C_2 et le sous-ensemble C_3 dédié aux conducteurs plus calmes. Par conséquent, chaque année, si le comportement de l'assuré se dégrade et atteint un sous-score de 25 ou moins, un malus (majoration de prime) sera accordé. En revanche, si le preneur d'assurance augmente son score à 75 ou plus, il recevra un bonus par une réduction de prime. Si le score d'un assuré se maintient entre 25 et 75, ce dernier continuera de payer la prime initialement fixée. Cependant, cette approche est trop directe et ne tient pas compte de la logique floue. La transition est brutale et un assuré qui passe d'un sous-score flou de 25 à 24 obtient une augmentation totale de prime, ce qui n'est pas en accord avec la théorie des ensembles flous.

La deuxième approche tient compte du mécanisme flou de déffuzification. Il ne faut

pas oublier que le score flou est un score numérique utilisé pour le système bonus/malus qui est construit via un processus de déffuzification. Dès lors, les scores compris entre 20 et 30 et entre 70 et 80 appartiennent respectivement aux sous-ensembles C_1 et C_2 et aux sous-ensembles C_2 et C_3 (voir figure 5.3.1 de droite). Pour ces tranches, il y a lieu d'ajuster le tarif progressivement, sinon l'approche floue perd de son sens. Dans le cas où l'appartenance totale (100%) d'un score à ses ensembles ne peut pas être déterminée, seule son appartenance aux sous-ensembles connus est considérée, comme si ceux-ci constituaient la totalité des ensembles auxquels cet objet appartient. Par exemple, le score 25 appartient à 25% au sous-ensemble C_1 et à 25% au sous-ensemble C_2 , la prime doit donc être impactée de manière égale (50%) par chacun de ces deux sous-ensembles, étant donné que C_1 et C_2 sont les seuls ensembles considérés. Pour le reste, l'approche 2 est semblable à l'approche 1.

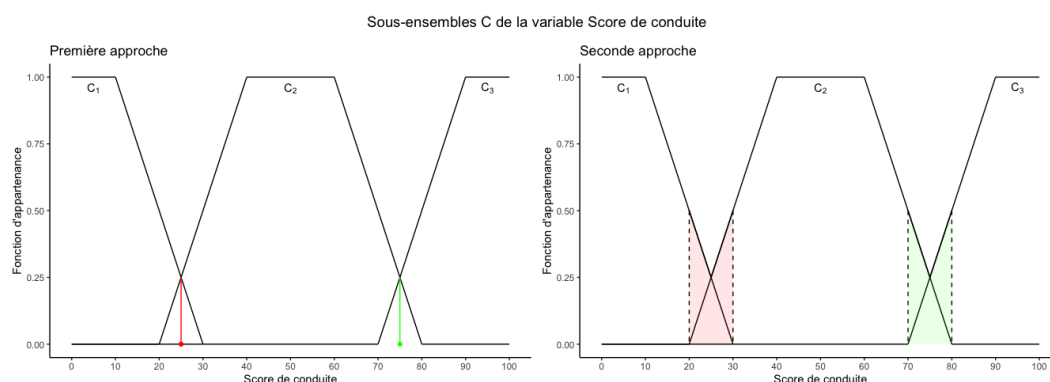


FIGURE 5.3.1 – Définition des frontières du système bonus-malus selon deux approches

La part de réduction ou d'augmentation de prime se réfère à la différence de sinistralité entre les différents types de conducteurs (calme, modéré, agressif). Si un assuré détient un score de 50 et améliore sa conduite en obtenant un score de 80 ou plus, l'assureur gagne en termes de sinistralité et lui fait en retour un cadeau. La réduction de prime sera égale à la différence entre la moyenne du taux de sinistres du groupe C_3 et celle du groupe C_2 . Si son score est compris entre 70 et 80, cette réduction sera revue à la baisse (car le score appartient aussi à C_2), proportionnellement à la part de son score appartenant à C_2 ¹. Au contraire, si cet assuré ne se comporte pas tel un conducteur prudent et que son score se dégrade à 20 ou moins, son augmentation de prime sera égale à la différence entre la moyenne du taux de sinistres du groupe C_2 et celle du groupe C_1 . Pour les scores compris entre 30 et 20, l'augmentation sera progressive et suivra le même raisonnement que celui décrit précédemment. Pour connaître cette différence de sinistralité, un GLM est réalisé sur les variables classiques et sur la variable du sous-score flou obtenu. Ensuite, pour chaque sous-ensemble flou C_1 , C_2 et C_3 , la fréquence de sinistres est prédite. La différence

1. Par exemple, un score de 73 appartient à 15% à C_3 et à 35% à C_2 . La prime associée doit donc être impactée à 70% ($35\% \times 2$) par le sous-ensemble C_2 et à 30% par C_3 ($15\% \times 2$). La diminution de prime pour un score égal à 73 sera donc égale à 30% de la différence de sinistralité entre les deux groupes.

de celle-ci selon les sous-ensembles flous déterminera la réduction/augmentation de prime. Toutefois, l'assureur doit garder une marge de sécurité. Dès lors, sur ces diminutions de prime, la compagnie peut garder un pourcentage pour se couvrir. Par raison d'égalité, ce même pourcentage est appliqué aux augmentations de prime. La sous-section 5.3.3 évoquera comment ce pourcentage peut jouer dans l'équilibre du portefeuille de l'assureur. Par ailleurs, seul le passage d'un sous-ensemble à un autre entraîne une variation de prime.

Au niveau du marketing, il est plus compréhensible pour le client que chaque sous-score impacte la prime de manière équivalente. Dès lors, les poids des augmentations/diminutions de prime pour chaque sous-score sont égaux. Par exemple, si un assuré obtient une augmentation de prime de 5% suite à un sous-score, mais reçoit une diminution de 5% suite à un des cinq autres sous-scores, l'impact final sur sa prime initiale sera alors de 0%. Le même procédé est appliqué pour l'ensemble des cinq sous-scores.

5.3.2 Exemple sur base de deux sous-scores

Afin d'illustrer ce produit télématique, la variation de prime d'un assuré est calculée en utilisant deux sous-scores de conduite choisis aléatoirement, à savoir les sous-scores "vitesse" et "conduite en heures de pointe". Les différents résultats obtenus se trouvent au tableau 5.3.1. La fréquence de sinistres diminue bien en passant du sous-ensemble C_1 (conducteurs agressifs, score plus petit ou égale à 30), aux sous-ensembles C_2 (conducteurs modérés, score compris entre 20 et 80) et C_3 (conducteurs calmes, score plus grand ou égale à 70). La marge de sécurité est posée arbitrairement à 50%. L'assureur la garde afin de se couvrir en cas de mauvaises surprises. Elle tient également compte de la marge bénéficiaire que souhaite se faire l'assureur sur l'aspect télématique.

Sous-score Vitesse	<=30	[20,80]	>=70
Fréquence moyenne de sinistres	0.04918	0.0456	0.0389
Variation de la fréquence de sinistres par rapport au sous-ensemble C_2	+ 0.0036	/	- 0.0068
Variation de la prime (50 %)	+ 0.0018	/	- 0.0034
Sous-score Conduite en heures de pointe	<=30	[20,80]	>=70
Fréquence moyenne de sinistres	0.052	0.0464	0.037
Variation de la fréquence de sinistres par rapport au sous-ensemble C_2	+ 0.0056	/	- 0.0097
Variation de la prime (50 %)	+ 0.0028	/	- 0.0048

TABLE 5.3.1 – Variation de la prime selon le système bonus/malus avec deux sous-scores

Soit un assuré qui détenait des sous-scores "vitesse" et "conduite en heures de pointe" égaux à 50. Un an plus tard, son sous-score "vitesse" est de 11.87 et son sous-score "conduite en heures de pointe" de 89.44. Celui-ci reçoit donc une augmentation de prime de 0.0018 liée à son score "vitesse" et une diminution de prime de 0.0048 liée à son bon comportement en heures de pointe. Sa diminution de prime finale est donc de 0.003. Chaque année, ce système doit être appliqué à tous les assurés en prenant l'ensemble des cinq sous-scores de conduite.

5.3.3 Équilibre du portefeuille

La conception d'un tel produit télématique peut être problématique d'un point de vue de l'équilibre du portefeuille. Il est difficile de prédire à l'avance comment les comportements des assurés vont se modifier et de déterminer l'impact de ces changements sur la sinistralité. Il est par conséquent important pour l'assureur de tester le maintien de l'équilibre de son portefeuille. Celui-ci doit notamment savoir comment réagir en cas de déséquilibre. Une des possibilités est donc de laisser la possibilité à l'assureur de garder de la marge pour réagir face à de quelconques surprises. Celle-ci a été supposée à 50%. Il est bien entendu possible de faire varier ce pourcentage en fonction du portefeuille en vigueur.

5.4 Perspective future de l'utilisation de l'approche fuzzy

Jusqu'à présent, l'assureur utilise les données qu'il possède afin de fixer des primes qui dépendent des caractéristiques analysées. L'assureur fait une offre à son client et ce dernier est libre de l'accepter. Une perspective future liée à cette théorie serait d'intégrer pour la première fois l'assuré dans un processus bilatéral de négociation. Si le client reçoit une grille lui permettant de traduire ses résultats numériques pour chacun de ses sous-scores en une modification de primes, il pourrait la remplir en tenant compte de ses propres caractéristiques. Comme c'est une approche floue, il est plus facile pour l'assuré de la compléter, car les variables sont plus faciles à comprendre. L'assuré serait donc capable de déduire, sur base de ses particularités, la construction définitive de sa prime et pourrait également faire une proposition de tarification à l'assureur. Une négociation doit ensuite être menée entre les deux parties. Cette approche permet un double gain. D'une part, l'assuré comprendrait le calcul réalisé par l'assureur et détiendrait donc des idées pour améliorer son comportement au volant. D'autre part, ce processus d'éducation permettrait à l'assuré d'obtenir une prime qui lui serait plus adaptée. L'assureur, quant à lui, augmenterait sa quantité d'information et son profit serait par conséquent moins mis en cause.

Une autre possibilité serait qu'après un temps d'observation, l'assureur propose à son client une variation de prime suivant les données qu'il possède. Si l'assuré s'engage à augmenter un score et qu'il respecte son engagement, il pourrait obtenir, en plus de sa variation de prime, un bonus supplémentaire. Cela responsabiliserait l'assuré dans le processus de tarification. Pour s'assurer que l'assuré suive le bon comportement, l'assureur procéderait à un suivi continu par la télématique.

Il est toutefois évident que cette approche est difficile à mettre en oeuvre. L'assureur doit contacter chacun de ses clients individuellement, ce qui représente un effort supplémentaire non négligeable. Une solution consisterait à ne seulement contacter que les personnes qui bénéficieraient le plus d'un contrat télématique. L'assuré doit également détenir des moyens pour arriver à ce genre de négociation. Cependant, c'est une vision futuriste et cela souligne l'intérêt que peut apporter l'approche floue dans l'assurance automobile.

Conclusion

L'objectif de ce mémoire était d'utiliser la théorie des ensembles flous afin de concevoir un produit télématique d'assurance automobile. Il a été prouvé que la théorie des probabilités classiques n'est pas adaptée pour le traitement et l'analyse des données télématiques et que le caractère flou de celles-ci doit être modélisé autrement. La théorie des ensembles flous développée par Zadeh (1965) a répondu à ce besoin. Celle-ci a permis de décrire le comportement routier des assurés répertoriés dans la base de données à l'appui de scores flous de conduite. Ces derniers ont servi à réviser la tarification de prime de chaque assuré.

Ce mémoire s'est appuyé sur une base de données simulée. Par conséquent, les résultats ont été donnés à titre d'exemple afin d'illustrer et d'appliquer la théorie floue. Il serait intéressant d'appliquer ce système à des données réelles afin d'étayer les résultats de cette analyse. De plus, il s'agit d'une étude actuarielle qui se base principalement sur la rationalité de l'assuré et qui ne tient pas compte de toutes les spécificités du monde réel. Selon ce mémoire, un assuré choisit un contrat télématique si ce dernier lui permet d'avoir la possibilité de réduire sa prime. Ce que pense l'assuré en ce qui concerne la sécurité et la protection de ses données n'a pas été pris en compte.

Cette approche floue permet de réfléchir à des perspectives futures dans le domaine de l'assurance automobile. Dans celles-ci, l'assuré pourrait être intégré dans le processus de négociation et l'assurance prendrait un nouveau tournant. L'idée que l'assuré entre dans un tel processus est dans la tendance actuelle. Par exemple, Corona Direct Assurances a lancé une assurance au kilomètre qui permet à l'assuré d'estimer le nombre de kilomètres qu'il envisage de rouler. Avec cette estimation, l'assureur adapte la prime de son assuré. Une application de la théorie floue à ce cas particulier permettrait d'améliorer l'ensemble du processus de tarification vers une prime encore plus adaptée, et de renforcer la relation bilatérale entre assureur et assuré.

Bibliographie

- Adamo, J. (1980, Novembre). Fuzzy decision trees. *Fuzzy Sets and Systems* 4(3), 207–219.
- Araujo, R., A. Igreja, R. de Castro, et R. E. Araujo (2012, Juin). Driving coach : A smartphone application to evaluate driving efficient patterns. In *2012 IEEE Intelligent Vehicles Symposium*, Alcal de Henares, Madrid, Spain, pp. 1005–1010. IEEE.
- Aumann, R. J. (1965, Août). Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications* 12(1), 1–12.
- Baoding Liu et Yian-Kui Liu (2002, Août). Expected value of fuzzy variable and fuzzy expected value models. *IEEE Transactions on Fuzzy Systems* 10(4), 445–450.
- Bertoluzza, C., N. Corral Blanco, et A. Salas (1995). On a new class of distances between fuzzy numbers. *Mathware & soft computing. 1995 Vol. 2 Núm. 2*.
- Bressand, C. (1993). À propos de la tarification de l’assurance automobile. *Économie & prévision* 108(2), 75–96.
- Butler, P., T. Butler, et L. L. Williams (1988). Sex-divided mileage, accident, and insurance cost data show that auto insurers overcharge most women.
- Castignani, G., R. Frank, et T. Engel (2013a, Octobre). Driver behavior profiling using smartphones. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, The Hague, Netherlands, pp. 552–557. IEEE.
- Castignani, G., R. Frank, et T. Engel (2013b, Octobre). An evaluation study of driver profiling fuzzy algorithms using smartphones. In *2013 21st IEEE International Conference on Network Protocols (ICNP)*, Goettingen, Germany, pp. 1–6. IEEE.
- Chawla, N. V., K. W. Bowyer, L. O. Hall, et W. P. Kegelmeyer (2002, Juin). SMOTE : Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16, 321–357.
- Chen, C., T. R. Razak, et J. M. Garibaldi (2020, Juillet). FuzzyR : An Extended Fuzzy Logic Toolbox for the R Programming Language. In *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Glasgow, United Kingdom, pp. 1–8. IEEE.
- Choquet, G. (1954). Theory of capacities. *Annales de l’institut Fourier* 5, 131–295.
- Colubi, A., R. Coppi, P. D’Urso, et M. A. Gil (2007). Statistics with fuzzy random variables. *Metron* 65(3), 277–303.
- de Campos Ibáñez, L. M. et A. González Muñoz (1989, Janvier). A subjective approach for ranking fuzzy numbers. *Fuzzy Sets and Systems* 29(2), 145–153.

- de Jong, P. et G. Z. Heller (2008, Février). *Generalized Linear Models for Insurance Data* (1 ed.). Cambridge University Press.
- De Runz, C., F. Blanchard, et M. Desjardin, Éric et Herbin (2008). Exploration d'un ensemble de quantités floues. In *Rencontres francophones sur la Logique Floue et ses Applications*, pp. 134–141. Cépaduès-Paris.
- de Wit, G. (1982, Octobre). Underwriting and uncertainty. *Insurance : Mathematics and Economics* 1(4), 277–285.
- Delgado, M., J. Verdegay, et M. Vila (1988, Avril). A procedure for ranking fuzzy numbers using fuzzy relations. *Fuzzy Sets and Systems* 26(1), 49–62.
- Denuit, M. et A. Charpentier (2005). *Mathématiques de l'assurance non-vie*. Collection "Economie et statistiques avancées". Paris : Economica.
- Denuit, M., D. Hainaut, et J. Trufin (2019). *Effective Statistical Learning Methods for Actuaries I : GLMs and Extensions*. Springer Actuarial. Cham : Springer International Publishing.
- Dubois, D. et H. Prade (1983, Septembre). Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences* 30(3), 183–224.
- Dubois, D. et H. Prade (1987, Décembre). The mean value of a fuzzy number. *Fuzzy Sets and Systems* 24(3), 279–300.
- Enderton, H. B. (1977). *Elements of set theory*. Academic press.
- Gil, M. A., M. A. Lubiano, M. Montenegro, et M. T. López (2002, Octobre). Least squares fitting of an affine function and strength of association for interval-valued data. *Metrika* 56(2), 97–111.
- González, A. (1990, Mars). A study of the ranking function approach through mean values. *Fuzzy Sets and Systems* 35(1), 29–41.
- González-Rodríguez, G., A. Colubi, et M. Ángeles Gil (2006, Novembre). A fuzzy representation of random variables : An operational tool in exploratory analysis and hypothesis testing. *Computational Statistics & Data Analysis* 51(1), 163–176.
- Goodfellow, I., Y. Bengio, et A. Courville (2016). *Deep learning*. Adaptive computation and machine learning. Cambridge, Massachusetts : The MIT Press.
- Grzegorzewski, P. (2008). Statistics with Vague Data and the Robustness to Data Representation. In D. Dubois, M. A. Lubiano, H. Prade, M. A. Gil, P. Grzegorzewski, et O. Hryniewicz (Eds.), *Soft Methods for Handling Variability and Imprecision*, Volume 48, pp. 100–107. Berlin, Heidelberg : Springer Berlin Heidelberg. Series Title : Advances in Soft Computing.
- Handel, P., I. Skog, J. Wahlstrom, F. Bonawiede, R. Welch, J. Ohlsson, et M. Ohlsson (2014). Insurance Telematics : Opportunities and Challenges with the Smartphone Solution. *IEEE Intelligent Transportation Systems Magazine* 6(4), 57–70.
- Horgby, P.-J., R. Lohse, et N.-A. Sittaro (1997). Fuzzy underwriting : an application of fuzzy logic to medical underwriting.
- Hosler, V. R. (1992). Actuarial research clearing house 1992 vol. 2 the application of fuzzy sets to group health underwriting.

- Kandel, A. et W. Byatt (1978). Fuzzy sets, fuzzy algebra, and fuzzy statistics. *Proceedings of the IEEE* 66(12), 1619–1639.
- Kruse, R. et K. D. Meyer (1987). Descriptive statistics with vague data. In *Statistics with Vague Data*, pp. 71–130. Dordrecht : Springer Netherlands.
- Le Breton, D. (2001, Décembre). Conduites à risque :. *Les cahiers de médiologie N° 12(2)*, 64–75.
- Lebart, L., M. Piron, et A. Morineau (2006). *Statistique exploratoire multidimensionnelle : Visualisation et inférences en fouille de données* (4e éd ed.). Sciences SUP. Paris : Dunod.
- Leekwijck, W. V. et E. E. Kerre (1999, Décembre). Defuzzification : criteria and classification. *Fuzzy Sets and Systems* 108(2), 159–178.
- Lemaire, J. (1990, April). Fuzzy insurance. *ASTIN Bulletin : The Journal of the IAA* 20(1), 33–55.
- Lerner, N., J. Jenness, J. Singer, S. Klauer, S. Lee, M. Donath, M. Manser, et N. Ward (2010). An exploration of vehicle-based monitoring of novice teen drivers : Final report. *Report number DOT HS 811*, 333.
- Leroy, D. (2014). Que sera l'industrie automobile, dans vingt ans? *Annales des Mines - Réalités industrielles Mai 2014(2)*, 23.
- Lubiano, M. A., M. A. Gil, M. López-Díaz, et M. T. López (2000, Mai). The α -mean squared dispersion associated with a fuzzy random variable. *Fuzzy Sets and Systems* 111(3), 307–317.
- Mamdani (1977, Décembre). Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis. *IEEE Transactions on Computers C-26(12)*, 1182–1191.
- Mamdani, E. et S. Assilian (1975, Janvier). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies* 7(1), 1–13.
- Mosmans, A., J.-C. Praet, et C. Dumont (2002, Juin). A decision support system for the budgeting of the Belgian health care system. *European Journal of Operational Research* 139(2), 449–460.
- Ohlsson, E. et B. Johansson (2010). *Non-life insurance pricing with generalized linear models*. EAA lecture notes. Heidelberg ; New York : Springer. OCLC : ocn471803012.
- Puri, M. L. et D. A. Ralescu (1986, Mars). Fuzzy random variables. *Journal of Mathematical Analysis and Applications* 114(2), 409–422.
- Runkler, T. (1997, Février). Selection of appropriate defuzzification methods using application specific properties. *IEEE Transactions on Fuzzy Systems* 5(1), 72–79.
- Schneider, M. et A. Kandel (1988, Juin). Properties of the fuzzy expected value and the fuzzy expected interval. *Fuzzy Sets and Systems* 26(3), 373–385.
- Schöner, H.-P. (2019). “How Good Is Good Enough ?” In Autonomous Driving. In J. Langheim (Ed.), *Electronic Components and Systems for Automotive Applications*, pp. 119–142. Cham : Springer International Publishing. Series Title : Lecture Notes in Mobility.
- Shapiro, A. F. (2007). An Overview of Insurance Uses of Fuzzy Logic. In S.-H. Chen, P. P. Wang, et T.-W. Kuo (Eds.), *Computational Intelligence in Economics and Finance*, pp. 25–61. Berlin, Heidelberg : Springer Berlin Heidelberg.

- Shashua, A., S. Shalev-Shwartz, et S. Shammah (2018). Implementing the rss model on nhtsa pre-crash scenarios. *tech. rep.*
- So, B., J.-P. Boucher, et E. A. Valdez (2021a). Cost-sensitive multi-class adaboost for understanding driving behavior based on telematics. *ASTIN Bulletin : The Journal of the IAA* 51(3), 719–751.
- So, B., J.-P. Boucher, et E. A. Valdez (2021b). Synthetic dataset generation of driver telematics. *Risks* 9(4), 58.
- Takagi, T. et M. Sugeno (1985, Janvier). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics SMC-15*(1), 116–132.
- Tselentis, D. I., G. Yannis, et E. I. Vlahogianni (2016). Innovative Insurance Schemes : Pay as/how You Drive. *Transportation Research Procedia* 14, 362–371.
- Vallin, R. W. (2013). *The elements of Cantor sets : with applications*. John Wiley & Sons.
- Wierman, M. J. (2010, Août). An Introduction to the Mathematics of Uncertainty. pp. 367.
- Yager, R. R. et L. A. Zadeh (Eds.) (1992). *An Introduction to Fuzzy Logic Applications in Intelligent Systems*. Boston, MA : Springer US.
- Zadeh, L. (1965, Juin). Fuzzy sets. *Information and Control* 8(3), 338–353.
- Zantema, J., D. H. van Amelsfort, M. C. J. Bliemer, et P. H. L. Bovy (2008, Janvier). Pay-as-You-Drive Strategies : Case Study of Safety and Accessibility Effects. *Transportation Research Record : Journal of the Transportation Research Board* 2078(1), 8–16.
- Zimmermann, H.-J. (2010, Mai). Fuzzy set theory : Fuzzy set theory. *Wiley Interdisciplinary Reviews : Computational Statistics* 2(3), 317–332.

Annexes

Annexe A

Présentation de la base de données

A.1 Provenance et brève explication de la base de données

Ce mémoire utilise une base de données regroupant un ensemble de 52 variables caractérisant un portefeuille d'assurance automobile de 100.000 polices (So et al., 2021b). Parmi ces variables, certaines sont associées à des risques traditionnels ou à la sinistralité des conducteurs et d'autres sont liées à la télématique. Ce sont des données synthétiques générées à l'aide de deux algorithmes d'apprentissage automatique populaires : le SMOTE étendu ("Synthetic Minority Oversampling Technique") et le réseau neuronal à anticipation. Le premier algorithme est utilisé afin de simuler les variables prédicteurs classiques et télématiques dans l'ensemble de données. Le second est quant à lui nécessaire pour générer le nombre de sinistres et le montant agrégé de ces derniers, qui sont les variables de réponse correspondantes. Pour plus d'information à ce sujet, l'article de revue de Chawla et al. (2002) pour le SMOTE et le livre de Goodfellow et al. (2016) pour le réseau neuronal à anticipation sont disponibles.

La simulation du portefeuille se base sur un ensemble de données réelles acquises auprès d'un assureur canadien, qui a lancé en 2013 un programme UBI (usage-based insurance) destiné à ses assurés. La période d'observation des données réelles s'étendait de 2013 à 2016 sur un portefeuille de plus 70.000 polices. Grâce à ces techniques, un portefeuille synthétique de 100.000 polices fut généré, dont 63 002 ont été utilisées pour l'entraînement et 36 998 pour le test final.

Les 52 variables explicatives sont divisées en onze variables traditionnelles, trente-neuf variables télématiques et deux réponses décrivant le nombre de sinistres déclarés et le montant cumulé des demandes de remboursement. Le tableau A.1.1 décrit chaque variable. Les variables "Pct.drive.xxx (mon/.../wed)" sont des variables de composition, ce qui signifie que la somme des sept variables (représentant les sept jours de la semaine) est égale à 1. Les variables "Pct.drive.wkday" et "Pct.drive.wkend" sont aussi des variables de composi-

tion. De plus, 95.72% des observations n'ont pas déclaré de sinistre, 4.06% en ont déclaré exactement un, seulement 0.20% en ont déclaré deux et enfin 0.01% ont réclamé trois sinistres.

Type	Variable	Description
Classique	Duration	Durée de la couverture d'assurance d'une police donnée en jours avec des valeurs comprises dans l'intervalle [22, 366].
	Insured.age	Âge du conducteur assuré, en années avec des valeurs comprises dans l'intervalle [16, 103].
	Insured.sex	Sexe du conducteur assuré (homme/femme)
	Car.age	Âge du véhicule, en années, avec des valeurs comprises dans l'intervalle [- 2, 20] ¹
	Marital	État civil (célibataire/marié)
	Car.use	Utilisation du véhicule : Privé, trajet domicile-travail, Agriculteur, Commercial
	Credit.score	Score de solvabilité du conducteur assuré
	Region	Type de région où vit le conducteur : rurale, urbaine
	Annual.miles.drive	Kilométrage annuel prévu déclaré par le conducteur
	Years.noclaims	Nombre d'années sans sinistre avec des valeurs comprises dans l'intervalle [0, 79].
	Territory	Localisation territoriale du véhicule. Cette variable a 55 libellés dans 11, 12, 13, . . . , 91.
Télématique	Annual.pct.driven	Pourcentage annualisé du temps passé sur la route
	Total.miles.driven	Distance totale parcourue en miles
	Pct.drive.xxx (mon/.../sun)	Pourcentage de conduite jour xxx de la semaine : lun/mar/. . . /dim
	Pct.drive.xhrs	Pourcentage de véhicules conduits dans un délai de x heures : 2 heures/3 heures/4 heures
	Pct.drive.xxx (wkday/wkend)	Pourcentage de véhicules conduits pendant xxx : jour/week-end
	Pct.drive.rushxx	Pourcentage de conduite pendant les heures de pointe de xx : am/pm
	Avgdays.week	Nombre moyen de jours d'utilisation par semaine
	Accel.xxmiles	Nombre d'accélération soudaines 6/8/9/. . . /14 mph/s par 1000 miles
	Brake.xxmiles	Nombre de freinages brusques 6/8/9/. . . /
	Left.turn.intensityxx	Nombre de virages à gauche par 1000 miles avec intensité 08/09/10/11/12
	Right.turn.intensityxx	Nombre de virages à droite par 1000 miles avec intensité 08/09/10/11/12
Réponses	NB_{Claim}	Nombre de sinistres pendant l'observation, compris entre 0 et 3.
	AMT_{Claim}	Montant cumulé des sinistres pendant l'observation

TABLE A.1.1 – Variables présentes dans la base de données

A.2 Division de la base de données

Lors de la construction d'un modèle, il est important que celui-ci reste général et soit efficace pour n'importe quel jeu de données. Le surentraînement, aussi appelé overfitting, doit être évité au maximum. Ce phénomène décrit un modèle qui s'applique bien aux valeurs du jeu de données utilisé pour l'entraînement de ce modèle, mais qui converge vers des valeurs incohérentes pour un autre jeu de données. Le modèle construit est donc inutile dans le monde réel. Dès lors, dans une bonne analyse, il est nécessaire de diviser la base de données en plusieurs parties. Le modèle est initialement ajusté sur un ensemble de données d'entraînement, afin de pratiquer à l'exploration, de sélectionner les variables et d'estimer les paramètres. Cette étape peut construire un ou plusieurs modèles. Le modèle ajusté va ensuite être utilisé pour prédire les réponses des observations de l'ensemble de données de validation. S'il existe plusieurs modèles, ce jeu de données est également utilisé pour comparer leurs performances et pour choisir le meilleur. La base de données utilisée dans ce mémoire a été divisée en 63% et 37% pour la base d'entraînement et de validation, respectivement. Chaque observation peut être utilisée pour une seule des bases de données. Cependant, bien qu'une observation peut être utilisée autant de fois que souhaité pour l'exploration, elle ne peut l'être qu'une seule fois pour la validation.

1. Les valeurs négatives désignent l'achat d'un modèle plus récent qui se réalise jusqu'à deux ans à l'avance.

A.3 Analyse exploratoire de certaines variables explicatives

Il est nécessaire d'analyser les différentes variables afin de comprendre la tendance générale des données. Afin d'optimiser les avantages de l'UBI pour les assureurs et leurs assurés, il devient crucial d'identifier les variables télématiques les plus significatives qui affectent réellement l'occurrence des accidents de la route. Les variables télématiques "Avg.days" et "Annual pct.driven" sont analysées.

A.3.1 Variable Avg.days

La moyenne de la variable "Avg.days" augmente avec la sinistralité (voir point noir sur la figure A.3.1). Par ailleurs, la variance de cette variable diminue quand le nombre de sinistres augmente. Cela montre qu'il existe une relation entre la sinistralité et le nombre de jours passés en moyenne par semaine sur les routes. Lorsque ce nombre est grand, l'assuré a plus de chance de causer un ou plusieurs dommages.

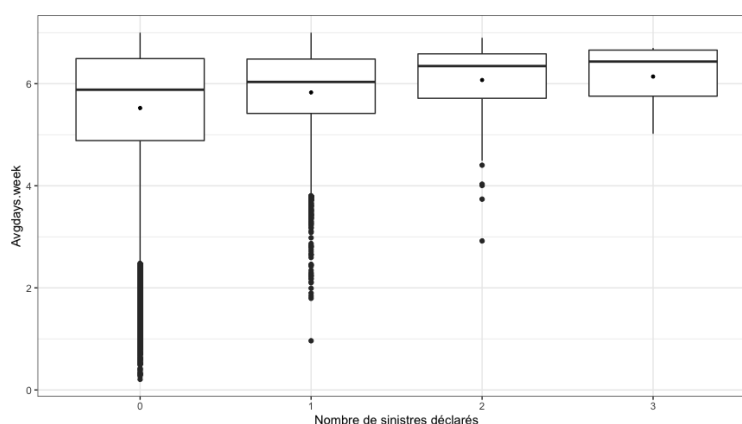


FIGURE A.3.1 – Boxplot de la variable Avgdays sur le nombre de sinistres

A.3.2 Variable Annual pct.driven

La moyenne de la variable "Annual pct.driven" augmente avec le nombre de sinistres (voir point noir sur le boxplot présent à la figure A.3.2). De plus, l'espace interquartile entre deux et trois sinistres diminue énormément. Cela prouve qu'il existe une relation entre le nombre de sinistres et le pourcentage de temps annuel passé sur la route. Lorsque ce pourcentage augmente, l'assuré a plus de chance d'être impliqué dans un sinistre.

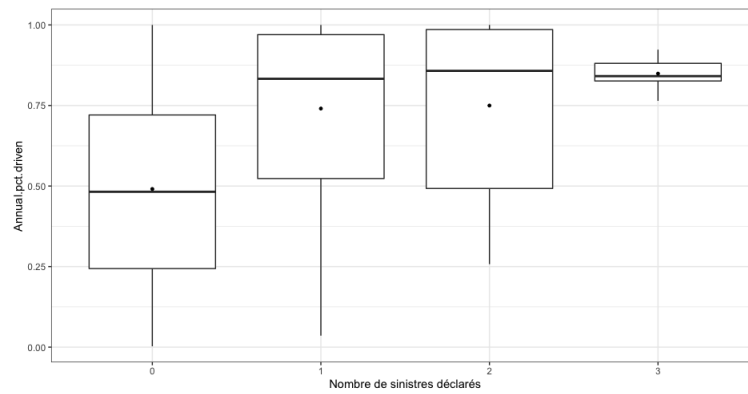


FIGURE A.3.2 – Boxplot de la variable Annual pct.driven en fonction du nombre de sinistres

Annexe B

Présentation du package FuzzyR

Afin de construire un système flou (voir chapitre 3), le package FuzzyR est utilisé. Celui-ci est une boîte à outils de la logique floue applicable au langage de programmation R. Un article reprenant toutes les caractéristiques de ce package fut écrit par [Chen et al. \(2020\)](#). FuzzyR est dérivé d'un ancien package nommé FuzzyToolkitUoN. Il a repris les différentes fonctions et en a ajouté de nombreuses autres. L'ancien package permettait entre autres de faire plusieurs choses, dont :

- Créer, évaluer et construire des fonctions d'appartenance, dont la majeure partie des formes couramment utilisées (triangulaires, trapézoïdales et gaussiennes) ont été définies et sont prêtes à être utilisées.
- Créer des systèmes d'inférence floue de style [Mamdani \(1977\)](#) avec des fonctions d'appartenance, des règles floues et des méthodes de défuzzification spécifiées par l'utilisateur.
- Permettre de visualiser la base de règles des systèmes d'inférence floue construits en utilisant des phrases IF-THEN faciles à comprendre.

Le nouveau package fuzzyR permet d'intégrer le système d'inférence floue de [Takagi et Sugeno \(1985\)](#) (TS). De plus, une interface utilisateur graphique (GUI) est présentée afin que tous les éléments d'un système flou, tels que les graphiques des fonctions d'appartenance, les règles floues et la défuzzification puissent être visualisés. Cet aspect rend ce package plus interactif.

Annexe C

Complément du chapitre 4

La fonction caractéristique $C(r)$ du sous-ensemble C est donnée par :

$$C(r) = \begin{cases} 0 & \text{si } r < 4.5 \text{ ou } r > 7 \\ \frac{r-4.5}{6-4.5} & \text{si } 4.5 \leq r \leq 6 \\ 1 & \text{si } 6 \leq r \leq 7 \end{cases}$$

Le tableau C.0.1 résume l'information nécessaire à la construction des fonctions de distribution et la figure C.0.1 illustre la fonction d'appartenance $C(r)$ ainsi que son complément et les fonctions de répartition F^* et F_* .

x	$r \leq x$	$\sup\{C(r) r \leq x\}$	$r > x$	$\inf\{1 - C(r) r > x\}$
0	0	0	$]0, \infty)$	0
1	$[0, 1]$	0	$]1, \infty)$	0
3	$[0, 3]$	0	$]3, \infty)$	0
5	$[0, 5]$	0.33	$]5, \infty)$	0
7	$[0, 7]$	1	$]7, \infty)$	1

TABLE C.0.1 – Quelques valeurs utiles à la construction des fonctions F^* et F_* pour le sous-ensemble C

Comme le noyau de C est représenté par $C(C) = [6, 7]$ (voir sous-section 2.3.4), $c^- = 6$ et $c^+ = 7$. Par conséquent, en utilisant les équations présentées en (4.2.5), les bornes de l'espérance du sous-ensemble flou C sont données par :

$$\begin{aligned} E_*(C) &= c^- - \int_{-\infty}^{c^-} C(x) dx \\ &= 6 - \int_{-\infty}^{4.5} 0 dx + \int_{4.5}^6 \frac{x - 4.5}{6 - 4.5} dx \\ &= 5.25 \end{aligned}$$

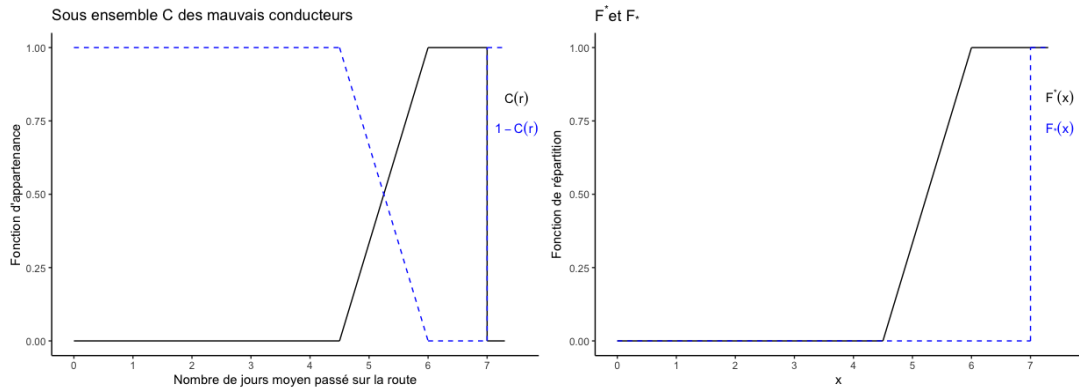


FIGURE C.0.1 – Sous-ensemble C représentant les conducteurs à plus hauts risques et les fonctions de distribution supérieure (F^*) et inférieure (F_*)

$$\begin{aligned}
 E^*(C) &= c^+ + \int_{c^+}^{+\infty} C(x) dx \\
 &= 7 + \int_7^{+\infty} 0 dx = 7
 \end{aligned}$$

Annexe D

Complément du chapitre 5

D.1 Analyse de corrélation et ACP

Variable	Description
Annual.pct.driven	Pourcentage annualisé du temps passé sur la route
Pct.drive.2hrs	Pourcentage de véhicules conduits dans un délai de 2 heures
Pct.drive.4hrs	Pourcentage de véhicules conduits dans un délai de 4 heures
Pct.drive.wkday	Pourcentage de véhicules conduits pendant la semaine
Pct.drive.rusham	Pourcentage de conduite pendant les heures de pointe de am
Pct.drive.rushpm	Pourcentage de conduite pendant les heures de pointe de pm
Avgdays.week	Nombre moyen de jours d'utilisation par semaine
Accel.06miles	Nombre d'accélération soudaines 6 mph/s par 1000 miles
Brake.06miles	Nombre de freinages brusques 6
Left.turn.intensity08	Nombre de virages à gauche par 1000 miles avec intensité 08
Right.turn.intensity08	Nombre de virages à droite par 1000 miles avec intensité 08

TABLE D.1.1 – Variables non corrélées de la base de données

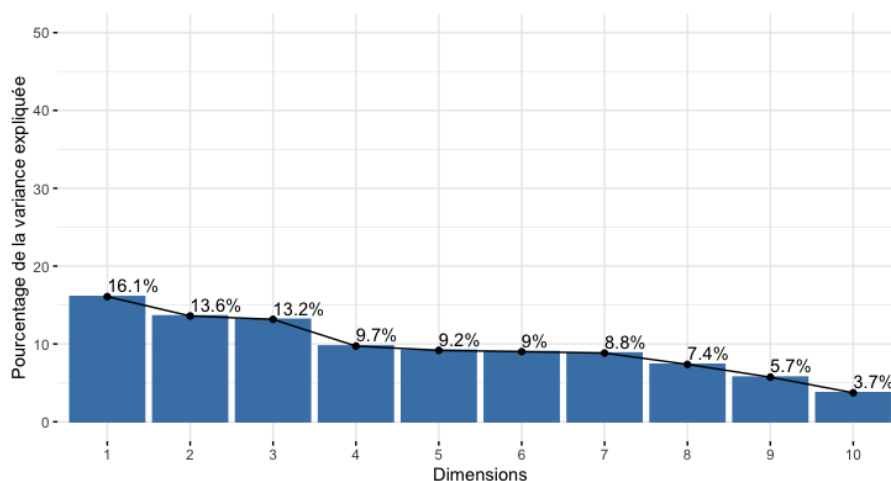


FIGURE D.1.1 – Scree plot

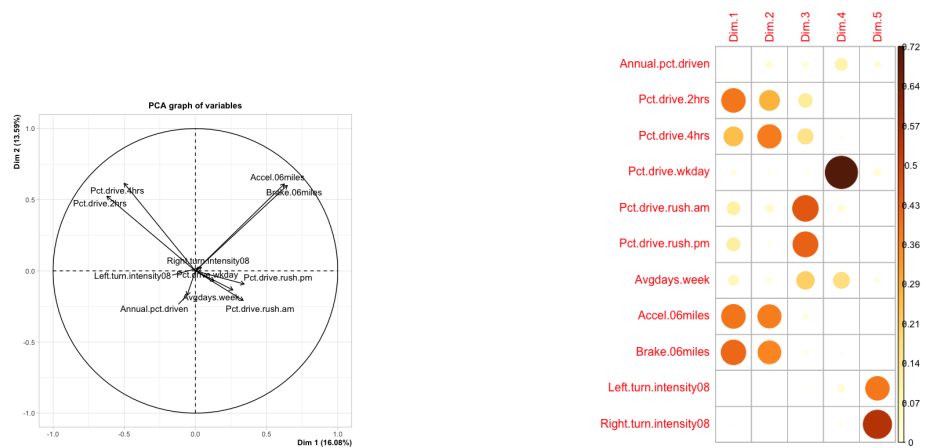


FIGURE D.1.2 – Cercle de corrélation (gauche) et qualité de représentation (droite)

D.2 Système d'inférence flou pour chaque sous-score

D.2.1 Sous-score 1 : vitesse

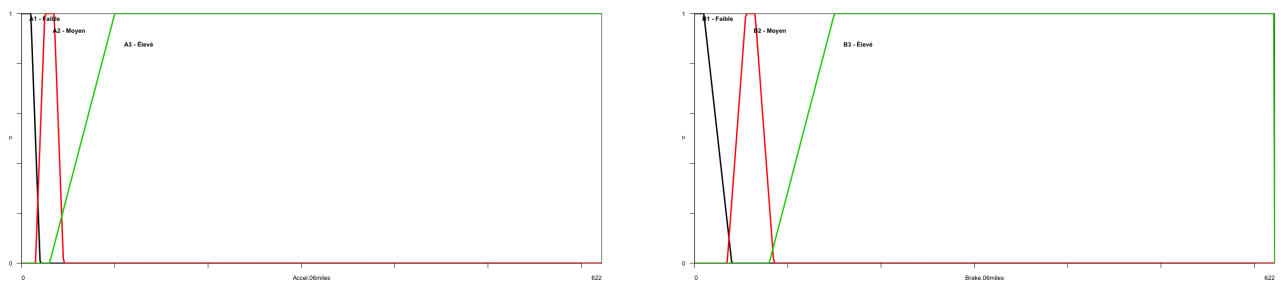


FIGURE D.2.1 – Sous-ensembles flous des variables "Accel.06mile" (gauche) et "Brake.06mile" (droite)

Variable	Nombre	Français	Nombre flou trapézoïdal
Accel.06mile	A_1	Faible	[0,0,10,20]
	A_2	Moyen	[15,25,35,45]
	A_3	Élevé	[30,100,622,622]
Brake.06mile	B_1	Faible	[0,0,10,40]
	B_2	Moyen	[35,55,65,85]
	B_3	Élevé	[80,150,621,622]

TABLE D.2.1 – Construction des nombres flous des variables "Accel.06mile" et "Brake.06mile"

Numéro de la règle	Accel.06mile	Et	Brake.06mile	Alors	Score de conduite
1	Faible (A_1)		Faible (B_1)		Calme (C_3)
2	Élevé (A_3)		Élevé (B_3)		Agressif (C_1)
3	Moyen (A_2)		Moyen (B_2)		Modéré (C_2)
4	Élevé (A_3)		Faible (B_1)		Modéré (C_2)
5	Faible (A_1)		Élevé (B_3)		Modéré (C_2)
6	Faible (A_1)		Moyen (B_2)		Modéré (C_2)
7	Moyen (A_2)		Faible (B_1)		Calme (C_3)
8	Élevé (A_3)		Moyen (B_2)		Modéré (C_2)
9	Moyen (A_2)		Élevé (B_3)		Agressif (C_1)

TABLE D.2.2 – Construction des règles floues pour le sous-score vitesse

D.2.2 Sous-score 2 : Heures de conduite

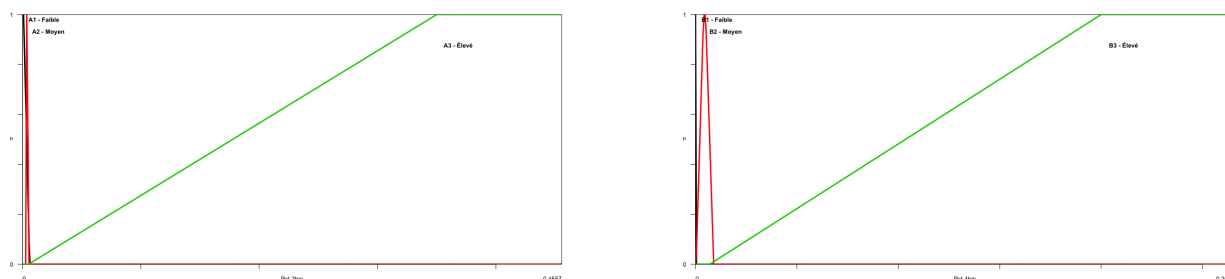


FIGURE D.2.2 – Sous-ensembles flous des variables "Pct.drive.2hrs" (gauche) et "Pct.drive.4hrs" (droite)

Variable	Nombre	Français	Nombre flou trapézoïdal
	A_1	Faible	$[0,0,0.001,0.006]$
Pct.drive.2hrs	A_2	Moyen	$[0.003,0.0035,0.004,0.00524]$
	A_3	Élevé	$[0.005,0.35,0.455743,0.455743]$
	B_1	Faible	$[0,0,0.0001,0.0005]$
Pct.drive.4hrs	B_2	Moyen	$[0.0003,0.004,0.005,0.009]$
	B_3	Élevé	$[0.007,0.2,0.2658871,0.2658871]$

TABLE D.2.3 – Construction des nombres flous des variables "Pct.drive.2hrs" et "Pct.drive.4hrs"

Numéro de la règle	Pct.drive.2hrs	Et	Pct.drive.4hrs	Alors	Score de conduite
1	Faible (A_1)		Faible (B_1)		Calme (C_3)
2	Élevé (A_3)		Élevé (B_3)		Agressif (C_1)
3	Moyen (A_2)		Moyen (B_2)		Modéré (C_2)
4	Élevé (A_3)		Faible (B_1)		Modéré (C_2)
5	Faible (A_1)		Élevé (B_3)		Modéré (C_2)
6	Faible (A_1)		Moyen (B_2)		Calme (C_3)
7	Moyen (A_2)		Faible (B_1)		Modéré (C_2)
8	Élevé (A_3)		Moyen (B_2)		Agressif (C_1)
9	Moyen (A_2)		Élevé (B_3)		Modéré (C_2)

TABLE D.2.4 – Construction des règles floues pour le sous-score pourcentage de véhicules conduits

D.2.3 Sous-score 3 : conduite en heures de pointe

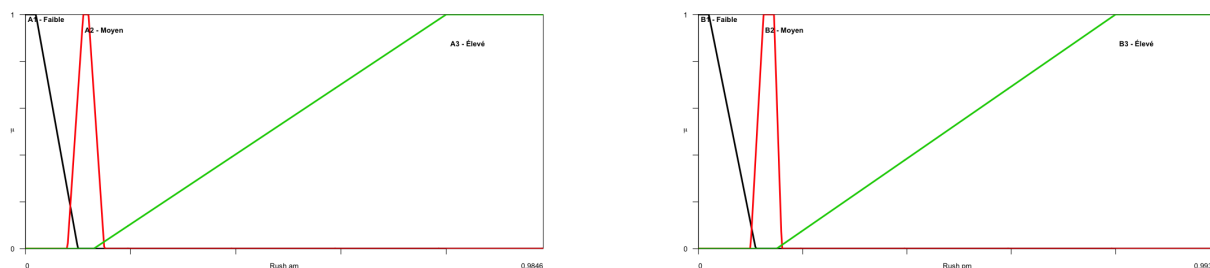


FIGURE D.2.3 – Sous-ensembles flous des variables "Pct.drive.rush.am" (gauche) et "Pct.drive.rush.pm" (droite)

Variable	Nombre	Français	Nombre flou trapézoïdal
Pct.drive.rush.am	A_1	Faible	$[0,0,0.02,0.10]$
	A_2	Moyen	$[0.08,0.11,0.12,0.15]$
	A_3	Élevé	$[0.13,0.8,0.9847, 0.9847]$
Pct.drive.rush.pm	B_1	Faible	$[0,0,0.02,0.11]$
	B_2	Moyen	$[0.10,0.125,0.145,0.16]$
	B_3	Élevé	$[0.15,0.8,0.9932,0.9932]$

TABLE D.2.5 – Construction des nombres flous des variables "Pct.drive.rush.am" et "Pct.drive.rush.pm"

Numéro de la règle	Pct.drive.rush.am	Et	Pct.drive.rush.pm	Alors	Score de conduite
1	Faible (A_1)		Faible (B_1)		Calme (C_3)
2	Élevé (A_3)		Élevé (B_3)		Agressif (C_1)
3	Moyen (A_2)		Moyen (B_2)		Modéré (C_2)
4	Élevé (A_3)		Faible (B_1)		Modéré (C_2)
5	Faible (A_1)		Élevé (B_3)		Modéré (C_2)
6	Faible (A_1)		Moyen (B_2)		Modéré (C_2)
7	Moyen (A_2)		Faible (B_1)		Calme (C_3)
8	Élevé (A_3)		Moyen (B_2)		Modéré (C_2)
9	Moyen (A_2)		Élevé (B_3)		Agressif (C_1)

TABLE D.2.6 – Construction des règles floues pour le sous-score conduite en heures de pointe

D.2.4 Sous-score 4 : temps passé sur la route

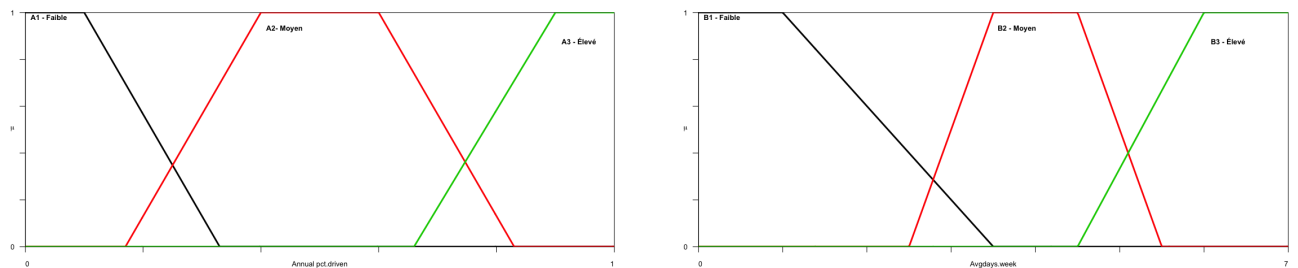


FIGURE D.2.4 – Sous-ensembles flous des variables "Annual pct.driven" (gauche) et "Avgdays.week" (droite)

Variable	Nombre	Français	Nombre flou trapézoïdal
Annual pct.driven	A_1	Faible	[0,0,0.1,0.33]
	A_2	Moyen	[0.17,0.4,0.6,0.83]
	A_3	Élevé	[0.66,0.9,1,1]
Avgdays.week	B_1	Faible	[0,0,1,3.5]
	B_2	Moyen	[2.5,3.5,4.5,5.5]
	B_3	Élevé	[4.5,6,7,7]

TABLE D.2.7 – Construction des nombres flous des variables "Annual pct.driven" et "Avgdays.week"

Numéro de la règle	Annual pct.driven	Et	Avgdays.week	Alors	Score de conduite
1	Faible (A_1)		Faible (B_1)		Calme (C_3)
2	Élevé (A_3)		Élevé (B_3)		Agressif (C_1)
3	Moyen (A_2)		Moyen (B_2)		Modéré (C_2)
4	Élevé (A_3)		Faible (B_1)		Modéré (C_2)
5	Faible (A_1)		Élevé (B_3)		Modéré (C_2)
6	Faible (A_1)		Moyen (B_2)		Calme (C_3)
7	Moyen (A_2)		Faible (B_1)		Modéré (C_2)
8	Élevé (A_3)		Moyen (B_2)		Agressif (C_1)
9	Moyen (A_2)		Élevé (B_3)		Modéré (C_2)

TABLE D.2.8 – Construction des règles floues pour le sous-score temps passé sur la route

D.2.5 Sous-score 5 : Prise des virages

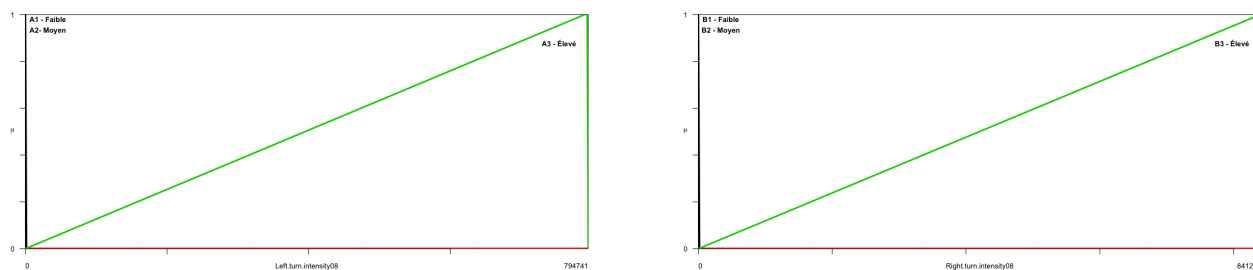


FIGURE D.2.5 – Sous-ensembles flous des variables "Left.turn.intensity08" (gauche) et "Right.turn.intensity08" (droite)

Variable	Nombre	Français	Nombre flou trapézoïdal
Left.turn.intensity08	A_1	Faible	[0,0,5,35]
	A_2	Moyen	[30,160,170,300]
	A_3	Élevé	[290,840000,841210,841210]
Right.turn.intensity08	B_1	Faible	[0,0,5,50]
	B_2	Moyen	[45,165,175,300]
	B_3	Élevé	[290,790000 ,794740 ,794740]

TABLE D.2.9 – Construction des nombres flous des variables "Left.turn.intensity08" et "Right.turn.intensity08"

Numéro de la règle	Left.turn.intensity08	Et	Right.turn.intensity08	Alors	Score de conduite
1	Faible (A_1)		Faible (B_1)		Calme (C_3)
2	Élevé (A_3)		Élevé (B_3)		Agressif (C_1)
3	Moyen (A_2)		Moyen (B_2)		Modéré (C_2)
4	Élevé (A_3)		Faible (B_1)		Modéré (C_2)
5	Faible (A_1)		Élevé (B_3)		Modéré (C_2)
6	Faible (A_1)		Moyen (B_2)		Calme (C_3)
7	Moyen (A_2)		Faible (B_1)		Modéré (C_2)
8	Élevé (A_3)		Moyen (B_2)		Agressif (C_1)
9	Moyen (A_2)		Élevé (B_3)		Modéré (C_2)

TABLE D.2.10 – Construction des règles floues pour le sous-score virages