

Titre : Approches statistiques semi-paramétriques pour données télématiques en assurance

Objectifs : La disponibilité de nouvelles données télématiques en assurance automobile fait en sorte qu'il est maintenant possible d'avoir un nouveau regard sur les activités d'assurance. La recherche scientifique sur ce sujet en est à ses débuts (Ayuso et al. 2016, Verbelen et al. 2016, Boucher et al. 2017) et les résultats sont prometteurs. Sous l'hypothèse que les nouvelles informations récoltées par les assureurs arriveront à mieux expliquer et comprendre le risque que les variables de segmentation usuelles (âge, sexe et état civil du conducteur, entre autres), nous développerons de nouvelles approches semi-paramétriques flexibles pour tarifier les produits d'assurance, ce qui permet d'estimer les probabilités d'accident et le coût de ces accidents.

Approche expérimentale ou théorique : De nouvelles données d'assurance réelles provenant d'un assureur seront utilisées et des données simulées serviront à notre analyse.

Méthodologie et procédure proposées: Pour développer une approche non paramétrique flexible, la théorie des modèles additifs généralisés pour paramètres de location-échelle sera une voie à étudier dans un premier temps. La flexibilité de ce modèle fait en sorte que des dépendances complexes entre variables aléatoires pourraient être considérées. En effet, les diverses couvertures ou chapitres d'assurance, les différents véhicules d'un même contrat d'assurance et les contrats d'assurance d'un même assuré observés à travers le temps créent une structure de dépendance hiérarchique complexe qu'il convient d'étudier pour mieux comprendre le risque. Une approche par effets aléatoires conjoints, ou par copule sera à étudier. Il pourrait être pertinent d'utiliser d'autres techniques d'apprentissage statistique adaptées à l'actuariat comme on peut les consulter par exemple chez Gao et al. (2018).

Connaissances actuelles : Les données télématiques servent à mieux cerner les habitudes de conduite de l'assuré et à mieux évaluer le style de conduite des assurées. Les approches semi-paramétriques de Boucher et al. (2017) ont montré le lien entre le temps d'exposition (la durée du contrat), le kilométrage parcouru et le nombre d'accidents. En comparant avec les techniques de tarification habituelles, des approches par réseau de neurones ou par analyse en composante principale (Ayuso et al. 2016, Verbelen et al. 2016, Gao et al., 2018) ont pu montrer l'avantage significatif de l'utilisation d'indicateurs de conduite (nombre d'accélération, vitesse moyenne, etc.).

Importance de la recherche: (1) Le projet de recherche proposé développera de nouveaux modèles statistiques adaptés aux caractéristiques des données télématiques dans lesquels une structure de dépendance hiérarchique sera incluse. (2) L'approche actuarielle qui sera développée devrait provoquer des changements importants dans les méthodes de tarification en assurance. (3) Nous verrons de quelle manière la télématique éliminera le besoin d'utiliser l'âge, le sexe, ou l'état civil du conducteur, variables de segmentation socialement discutables et maintenant proscrites d'utilisation dans plusieurs pays.

Références :

- Ayuso, M., Guillen, M. et Pérez-Marin, A.M. (2016). Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance. *Transportation Research Part C: Emerging Technologies*, **68**, 160-167.
- Boucher, J.-P., Côté, S. et Guillen, M. (2017). Exposure as duration and distance in telematics motor insurance using generalized additive models. *Risks* **5**(54).
- Gao, G., Meng, S. et Wüthrich, M.V. (2018). Claims frequency modeling using telematics car driving data. Technical Report.
- Verbelen, R., Antonio, K. et Claeskens, G. (2016). Unraveling the predictive power of telematics data in car insurance pricing. *SSRN ID: 2872112*.