

a) Problématique

Lorsqu'une compagnie d'assurances conclut un contrat avec un client, elle s'engage généralement à couvrir tous les sinistres qui surviendront pendant une période déterminée. Pour une catégorie de risque (automobile, habitation, etc.), les sinistres sont déclarés à l'assureur, parfois après un certain délai, puis payés après un laps de temps plus ou moins long. Les prestations à payer par une compagnie peuvent alors couvrir plusieurs années. À chaque instant, l'assureur doit donc conserver en réserve un certain montant de capital pour pouvoir faire face à ses obligations. À intervalles réguliers, la législation oblige l'assureur à faire la démonstration que ce montant en réserve est adéquat pour couvrir les risques liés aux engagements pris. Ces vérifications périodiques permettent la protection financière des assurés, des actionnaires et de la société en général et doivent être conduites minutieusement. Pour ce faire, il est nécessaire de compléter les dossiers de l'assureur, c'est-à-dire de prédire le montant et le moment de chacun des paiements qui seront effectués dans le futur, pour chacun des sinistres présents dans le portefeuille.

Ce problème peut être abordé selon plusieurs approches. Traditionnellement, les modèles procèdent de manière agrégée, c'est-à-dire en développant l'ensemble des coûts pour la totalité du portefeuille plutôt que le coût de chacun des sinistres individuellement. Ce point de vue a été étudié et critiqué depuis de nombreuses années (voir [EN02] et [WU08] pour une revue quasi exhaustive de la littérature sur ces modèles), en particulier sur le plan de la perte d'information que cette agrégation entraînait. Cependant, depuis le début du XXI^{ème} siècle, la multiplication des ressources informatiques et une disponibilité accrue des données dans les compagnies d'assurances ont permis l'émergence, chez les chercheurs en actuariat, d'approches alternatives basées sur la dynamique individuelle des sinistres. L'approche individuelle présente de nombreux avantages par rapport à celle collective car elle permet de mieux tenir compte de plusieurs éléments tels que l'hétérogénéité entre les sinistres, l'impact des clauses de réassurance, les caractéristiques variables des assurés, etc. Elle ouvre la porte à plusieurs développements qui étaient difficilement envisageables dans un cadre collectif tels que la détection automatique de situations inusitées (fraude, sinistres très coûteux, no touch claim, etc.) ou l'évaluation dynamique de la réserve en fonction de l'évolution de différents facteurs (présence d'un représentant légal, évaluation médicale, etc.). Plusieurs de ces objectifs peuvent, et doivent, être envisagés à la lumière des récents développements en matière d'apprentissage statistique et d'analyse de bases de données complexes et de grands volumes. L'utilisation de ces techniques en actuariat, plus particulièrement dans la modélisation des réserves, est très récente et la littérature scientifique (actuarielle) sur le sujet est quasi inexistante.

b) Objectifs de recherche

L'objectif principal est de développer des modèles individuels basés sur des techniques d'apprentissage statistique pour l'évaluation de la solvabilité d'un portefeuille d'assurance automobile. Plus spécifiquement, ce projet de recherche résultera en une connaissance approfondie de l'état actuel de la recherche en modélisation des réserves actuarielles, en de nouveaux modèles plus performants intégrant des outils statistiques et informatiques récents et en des recommandations quant à l'utilisation judicieuse de ce type d'approche pour l'évaluation de la solvabilité d'un portefeuille.

c) Méthodologie

Le projet suivra les étapes principales suivantes :

1. Réalisation d'une revue critique de la littérature sur l'utilisation de techniques d'apprentissage statistique en actuariat. En particulier, une étude de l'intérêt et de la faisabilité de certaines approches récemment présentées (par exemple dans [WU18]) sera réalisée.
2. Identification des particularités propres à l'évaluation individuelle du risque d'un portefeuille. À partir de modèles paramétriques individuels disponibles dans la littérature (par exemple dans [PI13]), nous analyserons l'impact d'intégrer certaines techniques d'apprentissage statistique dans le processus d'évaluation de la solvabilité.
3. Construction de modèles non paramétriques utilisant des outils issus de l'apprentissage statistique à partir des analyses réalisées à l'étape précédente.
4. Application à une ou plusieurs bases de données récentes et détaillées.

Pour les étapes 3 et 4, nous travaillerons à l'aide d'arbres de régression et de modèles adaptatifs (gradient boosting). Ces modèles seront construits en trois temps : (1) prédiction de la fréquence et de la sévérité des paiements futurs à l'aide de variables statiques (2) introduction de variables dynamiques et modélisation hiérarchique et (3) robustification et automatisation des modèles obtenus.

d) Contribution du projet à l'avancement des connaissances

Le projet de recherche proposé vise à mieux identifier, gérer et comprendre les risques d'assurance automobile au Canada. Tel que mentionné précédemment, la recherche portant sur la modélisation individuelle des réserves est récente mais se développe très rapidement. Les résultats obtenus dans ce projet permettront de poursuivre la recherche scientifique en actuariat en offrant de nouveaux modèles et de nouvelles techniques, en plus de contribuer à la réflexion entourant la gestion des risques par les différents acteurs.

e) Bibliographie

[EN02] England, P. et Verrall, R. (2002). *Stochastic claims reserving in general insurance*. British Actuarial Journal, 8(3), 443-544.

[PI13] Pigeon, M., Antonio, K. et Denuit, M. (2013). *Individual loss reserving with the multivariate skew normal framework*. ASTIN Bulletin, 43(3), 399-428.

[WU08] Wüthrich, M.V. et Merz, M. (2008). *Stochastic Claims Reserving Methods in Insurance*. Wiley Finance.

[WU18] Wüthrich, M.V. (2018). *Machine Learning in Individual Claims Reserving*. Scandinavian Actuarial Journal, 2018, 1-16.