

La modélisation des réserves actuarielles en utilisant de l'information agrégée a été largement développée par les chercheurs dans le passé et appliquée avec succès par les compagnies d'assurance. L'information agrégée est souvent présentée sous la forme d'un triangle de développement, dont chaque cellule représente la somme des paiements des réclamations pour une période donnée. En sommant l'ensemble des paiements, les caractéristiques individuelles de chaque réclamation sont perdues, alors qu'il serait probablement pertinent de les utiliser pour mieux estimer les réserves. C'est pourquoi il est aujourd'hui important de développer des modèles prenant en compte cette information plus détaillée.

Depuis le début des années 2000, plusieurs travaux ont été menés sur la modélisation des réserves en utilisant l'information au niveau individuel, la plupart proposant des modèles paramétriques, qui manquent souvent de flexibilité. Mes travaux effectués dans le cadre du mémoire de recherche proposent une approche non paramétrique à l'évaluation des réserves individuelles. Plus précisément, un algorithme d'apprentissage automatique construit en agrégeant séquentiellement des arbres de décision, appelé *gradient boosting*, a été ajusté aux données de réclamations d'une compagnie d'assurance. Le but ultime était de déterminer la distribution des pertes totales de la compagnie d'assurance. Un modèle *gradient boosting* est un modèle de prédiction : celui-ci prend en entrée les caractéristiques des réclamations et prédit les montants totaux remboursés pour chaque réclamation. Le modèle est entraîné sur un échantillon de réclamations qui se sont produites dans le passé.

Dans le développement de notre modèle, un traitement particulier a été fait pour considérer convenablement les données censurées. En effet, la vie d'une réclamation peut s'étaler sur plusieurs années, ce qui fait qu'à la date où la base de données est construite, certains dossiers sont encore ouverts, signifiant que le montant total remboursé n'est pas encore connu de manière exacte. Pour que le modèle de réserve estime correctement les montants futurs, ce dernier a besoin d'une valeur exacte dans l'étape d'entraînement de la technique d'apprentissage statistique. Pour corriger cette situation, nous avons opté pour une approche utilisant l'algorithme *Chain-Ladder*. En plus de cette considération, une adaptation de la base de donnée a été nécessaire pour tenir compte de l'effet longitudinal des données. En effet, la base de données de réserve décrit chacune des réclamations ainsi que leur évolution dans le temps.

Le modèle a été comparé avec deux approches largement utilisées en pratique : le *bootstrap Chain-Ladder* ainsi qu'un modèle linéaire généralisé prenant en compte l'information au niveau individuel. Ajusté aux données à ma disposition, le modèle *gradient boosting* donne des résultats qui surpassent ces deux modèles selon certains critères statistiques.