

Adaptation de l'algorithme CART pour la tarification des risques en assurance non-vie

Antoine Paglia^{*,***}, Martial Phélippe-Guinvarc'h^{**}, Philippe Lenca^{***,****}

^{*}EURO Institut d'actuariat EURIA, antoine.paglia@gmail.com

^{**}Actuaire, Direction des assurances Agricoles et Professionnelles, GROUPAMA SA
martialphelippeguinvarc'h@sfr.fr, martial.phelippe-guinvarc-h@groupama.com

^{***}Institut Télécom; Télécom Bretagne; UMR CNRS 3192 Lab-STICC,
philippe.lenca@telecom-bretagne.eu

^{****}Université européenne de Bretagne

Résumé. Les développements récents en tarification de l'assurance non-vie se concentrent majoritairement sur la maîtrise et l'amélioration des Modèles Linéaires Généralisés. Performants, ces modèles imposent cependant à la fois des contraintes sur la structure du risque modélisé et sur les interactions entre variables explicatives du risque. Ces restrictions peuvent conduire, dans certaines sous-populations d'assurés, à une estimation biaisée de la prime d'assurance. Les arbres de régression permettent de s'affranchir de ces contraintes et, de plus, augmentent la lisibilité des résultats de la tarification. Nous présentons une modification de l'algorithme CART pour prendre en compte les spécificités des données d'assurance non-vie. Nous comparons alors notre proposition aux modèles linéaires généralisés sur un portefeuille réel de véhicules. Notre proposition réduit les mesures d'erreur entre le risque mesuré et le risque modélisé, et permet ainsi une meilleure tarification.

1 Introduction

Les compagnies d'assurances utilisent quotidiennement des modèles statistiques pour évaluer les risques auxquels elles doivent faire face. L'objectif actuariel est d'estimer l'espérance de sinistre de chaque risque souscrit. Les trente dernières années ont été marquées par la sophistication des modèles de régression utilisés pour quantifier ces risques. La régression linéaire simple qui permettait de modéliser par une droite les variations d'une variable cible –le risque étudié–, a été remplacée à partir des années 1980 par les Modèles Linéaires Généralisés (McCullagh et Nelder, 1989), notés GLM par la suite. Ces modèles permettent à la fois de modéliser des comportements non linéaires et des distributions de résidus non gaussiens. Cela est particulièrement utile en assurance non-vie où les coûts des sinistres, quand ils se concrétisent, suivent une densité très asymétrique clairement non gaussienne. Ils ont permis d'améliorer nettement la qualité des modèles de prédiction du risque et sont aujourd'hui largement utilisés par les compagnies d'assurance. Cependant, bien que très sophistiqués, les modèles GLM présentent des limites en terme de modélisation des interactions entre variables explicatives et en