Dauphine | PSL UNIVERSITÉ PARIS

MIDO MATHÉMATIQUES ET INFORMATIQUE DE LA DÉCISION ET DES ORGANISATIONS

Master 1 - Mathématiques et Applications — Année 2020-2021

Mémoire Majeure Actuariat
–
Tarification en assurance IARD avec les GLM et les GAM

Encadrant	Christophe Dutang
Structure	Université Paris Dauphine
Courriel	dutang@ceremade.dauphine.fr
Données	Données réelles d'un assureur français

Contexte et objectif du projet

L'assurance est un contrat par lequel, moyennant le versement d'une prime dont le montant est fixé a priori (en début de période de couverture), l'assureur s'engage à indemniser l'assuré pendant toute la période de couverture (généralement un an). Cette prime doit refléter le risque associé au contrat. On renvoie à Charpentier et Denuit (2004a) sur la théorie du calcul des primes et à Charpentier et Denuit (2004b) pour les considérations économiques.

Pour chaque police d'assurance, la prime est fonction de variables dites de tarification permettant de segmenter la population en fonction de son risque. Il est usuel d'utiliser une approche fréquence/sévérité ou une approche indemnitaire pour modéliser le coût annuel d'une police d'assurance, voir cours Actuariat 1. Sur les données utilisées dans ce projet, on utilisera cette dernière approche car on ne dispose pas des montants individuels de sinistre.

Le but de ce projet est de proposer un tarificateur en se basant deux méthodes : les modèles linéaires généralisés (GLM) et les modèles additifs généralisés (GAM). Ces derniers sont une extension des GLM proposé par McCullagh et Nelder (1989) en considérant une approche non-paramétrique pour le prédicteur, voir Hastie et Tibshirani (1990). Plus précisément, pour un GLM, le prédicteur est une fonction linéaire des variables explicatives tandis que pour un GAM est non-paramétrique

$$\eta_i^{GLM} =< x_i, \beta >,\ \eta_i^{GAM} = \ \beta_0 + \sum_{j=1}^{p} f_j(x_{i,j}),$$

où x_i le vecteur de variables explicatives pour le $i$ème individu et f_j la $j$ème fonction lisse. Ce prédicteur permet de modéliser l'espérance de la variable réponse Y_i via une fonction dite "lien" $g(E(Y_i)) = \eta_i$ quand Y_i appartient à la famille exponentielle. Nous renvoyons aussi vers Efron et Hastie (2016) qui est disponible gratuitement en ligne mais moins complet que McCullagh et Nelder (1989) et Hastie et Tibshirani (1990).

Un second objectif sera, en plus de calculer une prime pure par police, de déterminer une commerciale intégrant une marge pour risque. Une approche par simulation sera réalisée pour juger de l'adéquation du chargement par rapport à la charge sinistre totale portefeuille.

Pour ce projet, les applications numériques seront à réaliser en R, R Core Team (2020), et on encourage les étudiants à étudier le livre de Charpentier (2014) sur ce sujet.

Guide pour le démarrage du projet

En utilisant les données du package R **CASdatasets**, Dutang et Charpentier (2019), les bases sinistre `freMPL3`, `freMPL4` d'assurance automobile va permettre de calibrer un GLM sur la sévérité et la fréquence. Comme déjà précisé, dans ces données, l'information sur la fréquence de sinistre est partielle car on sait juste si une police a eu au moins un sinistre sans en connaitre le nombre.

Les étapes suivantes devront être abordées :

1. Explorez les données `freMPL3` et `freMPL4` à l'aide des techniques exploratoires usuelles (statistiques descriptives, ACP, AFC).
2. Séparer les données d'apprentissage et de test à l'aide du code suivant
```
library(CASdatasets)
data(freMPL3) ; data(freMPL4)
freMPL34 <- rbind(freMPL3, freMPL4)
n <- NROW(freMPL34) ; p <- round(0.8*n)
set.seed(123) ; index.app <- sample(1:n, p)
freMPL34.app <- freMPL34[index.app, ] ; freMPL34.test <- freMPL34[-index.app, ]
```
3. Présenter les modèles linéaires généralisés en attachant une attention particulière aux lois considérées.
4. Modéliser la fréquence des sinistres sur la base `freMPL34.app` et la sévérité des sinistres sur la base `freMPL34.app` à l'aide des GLMs. Le choix des lois et la sélection des variables explicatives doivent être expliqués notamment à l'aide de statistique d'adéquation.
5. En déduire une prime pure pour les polices étudiées.
6. Présenter les modèles additifs généralisés, et en particulier leur méthode de calibration.
7. Modéliser la fréquence des sinistres sur la base `freMPL34.app` et la sévérité des sinistres sur la base `freMPL34.app` à l'aide des GAMs.
8. Sélectionner le meilleur modèle GAM en un sens qui sera à préciser.
9. Donnez deux tarifs en prime pure (un issu des GLMs l'autre des GAMs) pour les polices des données `freMPL34.test`. Comparez avec les valeurs observées de sinistre et conclure.
10. Par une analyse par simulation, déterminer une prime avec chargement (pour les deux approches GLM et GAM) afin que dans 99% des cas la somme des primes soit supérieure à la charge totale portefeuille des données `freMPL34.test`. On laisse le choix du principe de prime utilisé, voir cours Actuariat 1. Enfin calculez les primes avec chargement sur les données `freMPL34.test` et comparer avec les observations

Références

Charpentier, A. et Denuit, M. (2004a). Mathématiques de l'assurance non vie. T. 1. Economica.
— (2004b). Mathématiques de l'assurance non vie. T. 2. Economica.
Charpentier, A., éd. (2014). Computational Actuarial Science with R. Chapman et Hall-CRC.
Dutang, C. et Charpentier, A. (2019). CASdatasets : Insurance datasets.
Efron, B. et Hastie, T. J. (2016). Computer Age Statistical Inference. Cambridge University Press.
Hastie, T. J. et Tibshirani, R. J. (1990). Generalized Additive Models. Chapman et Hall.
McCullagh, P. et Nelder, J. A. (1989). Generalized Linear Models. 2nd. Chapman et Hall.
R Core Team (2020). R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL : https://www.R-project.org/.