

TER Magistère 1ère année

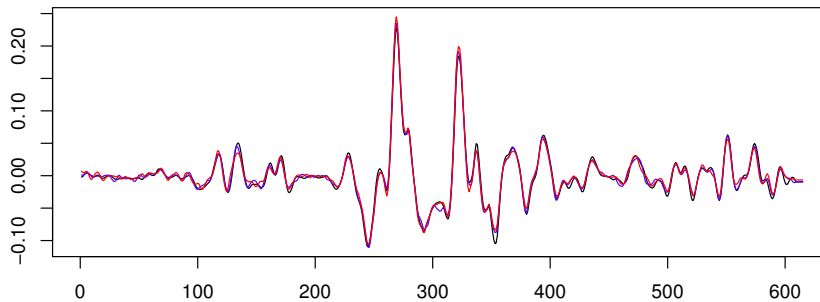
Prédire le grade de stéatose hépatique sans biopsie du foie

Contexte

La stéatose hépatique est une accumulation de graisse dans les cellules du foie. Cette maladie, aux conséquences graves, est due à une consommation excessive d'alcool ou à une alimentation trop riche en graisses et sucres. Elle touche une part de plus en plus importante de la population.

Aujourd'hui, le grade de stéatose est déterminé à partir d'une biopsie du foie. DIAFIR, start-up rennais, propose une méthode innovante basée sur la mesure spectrale d'un échantillon de sang associée à un modèle statistique qui suppose que la forme du spectre permet de prédire le grade de stéatose. Cette méthode évite la biopsie. Mais le modèle statistique a besoin d'être amélioré.

La figure ci-dessous montre des exemples de spectres pour 3 individus. Pour chacun de ces individus, on connaît le grade de stéatose. En pratique, on dispose de données pour 70 individus.



Projet

La régression linéaire multivariée permet de prédire le grade de stéatose à partir des données spectrales (variables explicatives) par la formule

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

où chaque X_j correspond par exemple à un point sur la courbe.

Ici, les variables sont considérées comme des courbes et on sait, biologiquement, que l'information concernant Y est localisée sur une partie de la courbe mais on ne sait pas localiser cette partie commune à tous les individus.

L'objectif de ce projet est de développer une méthode de régression qui permette d'analyser ce problème. Pour cela on fabriquera une nouvelle famille de variables explicatives construites à partir des courbes. L'idée est d'approcher les courbes par des polynômes du second degré sur une famille d'intervalles de longueur et de localisation différentes. Les coefficients des polynômes permettent alors de caractériser la courbe à différentes échelles (longueur de l'intervalle) et à différents endroits (position de l'intervalle). Dans un second temps, on utilisera toutes ces variables dans un modèle de régression linéaire et on fera l'estimation par des méthodes classiques permettant en particulier de mettre en évidence les variables (et donc les zones) les plus influentes.

Contacts

Bernard Delyon (bernard.delyon@univ-rennes1.fr)

Valérie Monbet (valerie.monbet@univ-rennes1.fr)