

Évaluation de la régression bornée

Thierry Foucart

UMR 6086, Université de Poitiers, S P 2 M I, bd 3 téléport 2
BP 179, 86960 Futuroscope, Cedex FRANCE

Résumé. Le modèle linéaire est très fréquemment utilisé en statistique et particulièrement dans les secteurs de l'assurance, de la banque et du marketing. Il permet de déterminer les variables explicatives qui interviennent dans le risque mesuré chez les assurés et dans les choix effectués par la clientèle. Le problème considéré dans cet article apparaît lorsque ces variables sont liées statistiquement, par exemple le revenu et la catégorie socioprofessionnelle. Les estimations données par le critère des moindres carrés ordinaires deviennent alors instables et peuvent prendre des valeurs en contradiction avec les valeurs réelles. Il existe de nombreuses méthodes adaptées à ce type de données. Nous proposons ici d'évaluer l'efficacité de la régression bornée en procédant par simulation. Les résultats sont clairs : le gain en précision et en stabilité des coefficients de régression est impressionnant.

1 Introduction

Le modèle linéaire est une des méthodes statistiques les plus employées dans les sciences de l'homme et de la société. Il donne en effet une réponse à la question récurrente de l'effet propre d'une variable sur une autre. En assurance automobile par exemple, la question pourrait être : l'âge du conducteur joue-t-il un rôle dans le risque d'accident indépendamment des autres facteurs ? Ce risque dépend-il de son sexe toutes choses égales par ailleurs ? Pour répondre à ces questions, on effectue la régression du risque par les facteurs d'accident, et on étudie chacun des coefficients de régression de l'âge et du sexe : la réponse est considérée comme positive si ce coefficient est significativement non nul.

L'hypothèse « toutes choses égales par ailleurs », formalisée par le choix des variables explicatives du modèle linéaire, est toutefois très discutée depuis fort longtemps parce qu'elle ouvre la porte à des abus flagrants (Simiand, 1932). Sa formalisation demande beaucoup de précautions pour éviter des contradictions internes. Ces dernières se manifestent au plan mathématique par une relation linéaire exacte entre les variables explicatives. Dans ce dernier cas, l'analyse statistique est impossible, la matrice de corrélation n'étant pas inversible.

Ces contradictions ne sont pas toujours totales. Il existe des situations dans lesquelles les variables explicatives ne sont pas liées au sens linéaire du terme (il n'existe pas de combinaison linéaire strictement égale à 0), mais le sont au sens statistique (il existe une combinaison linéaire « presque » égale à 0). L'estimateur des moindres carrés ordinaires devient alors peu précis, et on est amené à utiliser d'autres estimateurs dont les plus classiques sont ceux de la