

La régression inverse par tranches ou méthode SIR : présentation générale

Saracco Jérôme¹, Larramendy Irène¹ et Aragon Yves²

¹ Laboratoire de Probabilités et Statistique,
Département de mathématiques, case courrier 051,
Université Montpellier II,
Place Eugène Bataillon,
34 095 Montpellier Cédex 5.

Email : saracco@stat.math.univ-montp2.fr
larramen@stat.math.univ-montp2.fr

² GREMAQ,
Université Toulouse I,
Manufacture des Tabacs,
21 Allée de Brienne,
31 000 Toulouse.
Email : aragon@cict.fr

La régression inverse par tranches ou méthode SIR (Sliced Inverse Regression) est une méthode de régression semiparamétrique reposant sur un argument géométrique. Elle a été introduite par Li (1991). Dans le modèle, la variable à expliquer peut être unidimensionnelle ou multidimensionnelle, la variable explicative doit être multidimensionnelle. Au contraire des autres méthodes de régression semiparamétrique, elle ne requiert que des temps de calcul informatique très courts.

Nous donnons dans ce document une présentation générale des méthodes SIR univariées et multivariées. L'implémentation informatique de ces méthodes, réalisée dans le logiciel Splus, est présentée dans un second article de cette même revue et des illustrations de ces méthodes y sont proposées, voir Saracco (1998).

1 Introduction

Un modèle de régression paramétrique décrit les relations entre une variable à expliquer y à valeurs dans \mathbb{R}^q ($q \geq 1$) et une variable explicative x à valeurs dans \mathbb{R}^p ($p \geq 1$) de la forme :

$$y = f_{\theta}(x) + \epsilon,$$

où f_{θ} appartient à une famille de fonctions paramétrées par θ , vecteur de paramètres réels et où ϵ est un terme d'erreur aléatoire. Dans un tel modèle, l'objectif est l'estimation du paramètre θ . Les techniques d'estimation paramétrique (citons par exemple les méthodes du maximum de vraisemblance ou des moindres carrés) sont efficaces quand la famille de f_{θ} est correctement spécifiée.