

REVUE DE STATISTIQUE APPLIQUÉE

L. FERRÉ

A. F. YAO

Un critère de choix de la dimension dans la méthode SIR II

Revue de statistique appliquée, tome 47, n° 4 (1999), p. 33-46

http://www.numdam.org/item?id=RSA_1999__47_4_33_0

© Société française de statistique, 1999, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

UN CRITÈRE DE CHOIX DE LA DIMENSION DANS LA MÉTHODE SIR II

L. Ferré*, A.F. Yao**

* *Groupe de recherche en Informatique et Mathématiques du Mirail, département
de Mathématiques et Informatique*

*Université Toulouse Le Mirail, 5, Allées Antonio Machado,
31058 Toulouse Cedex, France
e-mail : loferre@univ-tlse2.fr*

** *Laboratoire de Statistique et Probabilités, URM R C 5583*

*Université Paul Sabatier, 118, route de Narbonne
31062 Toulouse Cedex, France
e-mail : yao@cict.fr*

RÉSUMÉ

La méthode «Sliced Inverse Regression» (SIR) a été introduite par Li (1991) dans les problèmes de régressions multidimensionnelles pour réduire la dimension par projection sur un sous espace, S_K , de l'espace initial. Cependant, dans la version initiale, certaines dimensions pertinentes peuvent ne pas être prises en compte, ce qui a conduit à une méthode plus adéquate, la SIR II. La difficulté principale est la détermination de la dimension du modèle final et, en adaptant une démarche efficace pour la SIR à la méthode SIR II, nous proposons un critère pour choisir la dimension, basé sur la qualité de l'estimation de l'espace S_K . Un estimateur de ce critère est étudié sur un plan théorique, mais surtout sur des simulations et des données réelles.

Mots-clés : *projecteur propre, loi elliptique symétrique, modèle de régression, choix de dimension.*

ABSTRACT

Sliced Inverse Regression is a method for reducing the dimensionality in multivariate non parametric regression problems. Unfortunately, some relevant directions may not be taken into account. Alternative methods have been proposed to overcome this problem and SIR II is one of them. However, the problem of determining the dimension of the model still remains. We propose here the suitable version of a criterion introduced in SIR. An estimate of this criterion is introduced and studied on both simulations and a real data set.

Keywords : *eigenprojection, elliptically symmetric distribution, general regression model, choice of the dimensionality.*

1. Introduction

Considérons le problème de la régression d'une variable explicative y sur une variable p -dimensionnelle x . Il est bien connu que les vitesses de convergence obtenues par des méthodes non-paramétriques sont pénalisées par la dimension p et nombreuse est la littérature traitant du sujet. Par exemple, la méthode de «projection pursuit regression» (Friedman and Stuetzle, 1981), la méthode ACE (Brieman and Friedman, 1985), la méthode GAM (Hastie and Tibshirani, 1986), la méthode ADE (Härdle and Stoker, 1989) visent à substituer à la régression multiple, des régressions unidimensionnelles, pour obtenir des vitesses de convergence comparables à celles des techniques univariées. Hormis ADE, toutes présentent l'inconvénient de reposer sur des algorithmes itératifs complexes.

La méthode «Sliced Inverse Regression» (SIR) proposée par Li (1991) permet d'aborder le problème sans avoir recours à de tels algorithmes. En supposant que la régression de y en x ne dépend que de la projection de x sur un sous-espace S_K de dimension K , ce dernier est estimé à partir de la simple diagonalisation d'une matrice. On utilisera la modélisation suivante

$$y = f(\beta_1^t x, \dots, \beta_K^t x, \epsilon), 1 \leq K \leq p, \quad (1)$$

où les vecteurs β_i , $i = 1, \dots, K$, sont indépendants et engendrent S_K , ϵ étant une variable aléatoire réelle indépendante de x . Ainsi, si K est raisonnablement petit, on pourra convenablement estimer f au moyen des méthodes non-paramétriques classiques.

Sous une condition vérifiée, en particulier si x suit une loi elliptique symétrique, mais qui est approximativement vérifiée par la plupart des lois multidimensionnelles (Hall et Li, 1993), Li montre que le sous-espace associé aux valeurs propres non nulles de la matrice $\Sigma^{-1} \text{var}(E(x/y))$ (où $\Sigma = \text{var}(x)$) est inclus dans l'espace S_K , appelé espace E.D.R (Effective Dimension Reduction). Nous appellerons, par la suite, SIR I cette méthode.

Cependant, il peut arriver que SIR I se montre inefficace pour estimer l'espace E.D.R dans son entier. En effet, si la projection de x dans certaines directions est de loi symétrique et si la régression de y par rapport à cette projection l'est également, alors, la matrice $\text{var}(E(x/y))$ sera dégénérée dans ces directions et celles-ci ne pourront être estimées correctement. Pour remédier à ce problème, deux stratégies ont été proposées. La première, la méthode SIR II (Li, 1991), repose sur l'utilisation de moments d'ordres supérieurs. Cette méthode sera développée par la suite. La seconde, méthode PHD («Principal Hessian Direction» Li, 1992), s'appuie sur l'estimation d'une matrice «asymétrique», $E(yx^t x)$, qui dans le cas normal se déduit de l'estimation de l'espérance de la matrice Hessienne de f .

Cependant, l'une ou l'autre des méthodes fournit une liste d'estimateurs de S_K sans pour autant permettre, à priori, de préjuger du meilleur : la dimension K est inconnue et se doit d'être estimée. Pour les méthodes SIR I et PHD, Ferré (1998) présente un critère de choix de modèle visant à optimiser l'estimation de l'espace E.D.R. Ce critère est étendu dans Ferré (1997) à la variante de SIR proposée par Hsing et Carroll (1992). L'idée est de déterminer K en étudiant le comportement de

la fonction de perte définie, pour $q = 1, \dots, K$, par

$$R(q) = 1 - \frac{1}{q} E(\text{tr}(\widehat{\Pi}_q \Pi_q)),$$

où Π_q et $\widehat{\Pi}_q$ représentent respectivement des projecteurs sur un sous-espace de dimension q de l'espace E.D.R et sur son estimateur. Ces fonctions mesurent les «écarts» respectifs entre les directions «idéales» et les directions observées. Ainsi, on ne retiendra que les directions pour lesquelles $R(q)$ est proche de zéro. Nous proposons ici de développer une approche semblable pour la méthode SIR II. Notons que, dans ce cadre, le problème de la détermination de K a été abordé au moyen de tests d'hypothèses par Schott (1994) et par Kötter (1996) au moyen d'intervalles de confiance autour du «pourcentage de variance reproduite» par les directions, sans toutefois conduire à une solution satisfaisante.

Dans la section 2, nous rappelons les principes de la méthode SIR II et définissons notre critère. La troisième partie est consacrée aux applications. Une étude au travers de simulations est menée pour évaluer la capacité du critère à retenir une dimension correcte. Puis, la méthode SIR II est mise en œuvre sur des données psychologiques issues de Zuroff (1994). L'objectif est d'expliquer l'intensité d'une relation amoureuse par sept variables décrivant des aspects de la personnalité des individus. Le résultat obtenu montre que trois combinaisons linéaires des sept variables dépendantes suffisent à l'explication de l'intensité amoureuse.

2. Modèle et choix de dimension

2.1. La méthode SIR II

Pour fixer les idées sur l'intérêt de la méthode SIR II, considérons l'exemple suivant (Li, 1991) : Soit $(x_1, x_2) \sim \mathcal{N}(0, I_2)$, et $y = x_1^2$; du fait de l'indépendance entre x_1 et x_2 , $E(x_2/y) = 0$, et de part la symétrie, $E(x_1/y) = 0$. Ainsi, $\text{var}(E(x/y)) = 0$ et l'estimation de $\beta = (1, 0)^t$ est alors impossible par SIR I. En revanche, si on considère les moments d'ordre 2 conditionnels, on a : $\text{var}(x_1/y) = E(x_1^2/y) = y$ et $\text{var}(x_2/y) = 1$, et β s'obtient en utilisant la matrice de variance-covariance conditionnelle; ceci est en fait l'idée de SIR II.

D'une façon générale, si P est le projecteur sur l'espace S'_K , engendré par $(\eta_1, \dots, \eta_K) = (\Sigma^{1/2}\beta_1, \dots, \Sigma^{1/2}\beta_K)$, et Q le projecteur sur son orthogonal, et si x suit une loi elliptique symétrique, on a, en posant $z = \Sigma^{-1/2}x$, la relation suivante :

$$\text{var}(z/y) = \omega_y Q - P \text{var}(z/y) P$$

(voir Cook and Weisberg, 1991), où ω_y est une variable aléatoire réelle dépendant de y et de la loi de x . Cette variable vérifie également $E(\omega_y) = 1$, et ω_y est identiquement égale à 1 si x suit une loi normale. Ainsi, tout vecteur v orthogonal à l'espace S'_K est vecteur propre de la matrice $\text{var}(z/y)$, associé à la valeur propre ω_y . En fait, dans Li (1991), il est montré que, déterminer l'espace S'_K , revient à chercher les directions

η_1, \dots, η_K maximisant, sous contrainte d'orthogonalité, la distance :

$$E (\|(\text{var}(z/y) - E(\text{var}(z/y)))\eta\|^2)$$

et donc à diagonaliser la matrice :

$$E (\text{var}(z/y) - E(\text{var}(z/y)))^2$$

dont les vecteurs propres, engendrant S'_K , sont les (η_1, \dots, η_K) . Une fois les vecteurs η_1, \dots, η_K obtenus, on en déduit l'espace E.D.R qui est engendré par les vecteurs $b_k = \Sigma^{-1/2}\eta_k$. Les vecteurs b_k peuvent s'obtenir directement comme vecteurs propres de la matrice :

$$\Sigma^{-1}E [(\text{var}(x/y) - E(\text{var}(x/y)))\Sigma^{-1}(\text{var}(x/y) - E(\text{var}(x/y)))] .$$

Comme pour la SIR I, et pour simplifier les méthodes d'estimation, la matrice ci-dessus sera remplacée par une matrice $\Sigma^{-1}W$, telle que W soit obtenu par tranchage du support de y . En effet, soit H le nombre total de tranches et I_h la $h^{\text{ième}}$ tranche, $h = 1, \dots, H$, $p_h = P(y \in I_h)$, $C_h = \text{var}(x/y \in I_h)$, $\Delta = E(C_h)$, $K_h = C_h - \Delta$ et n_h le nombre d'observations dans I_h , on pose

$$W = \sum_{h=1}^H p_h K_h \Sigma^{-1} K_h.$$

On suppose maintenant que la matrice $\Sigma^{-1}W$ possède K valeurs propres simples, $(\lambda_i)_{i=1, \dots, K}$, ordonnées dans l'ordre décroissant. On peut alors montrer que l'espace E.D.R contient le sous-espace propre associé à $(\lambda_1 \dots \lambda_K)$. On notera que, sous le modèle (1), $\Sigma^{-1}W$ a également une valeur propre multiple associée à l'orthogonal de l'espace E.D.R notée λ_{K+1} . Contrairement à la SIR I, cette valeur propre n'est pas nulle.

Un estimateur de W est donné par :

$$\widehat{W} = \sum_{h=1}^H \frac{n_h}{n} \widehat{K}_h \widehat{\Sigma}^{-1} \widehat{K}_h,$$

où $\widehat{K}_h = \widehat{C}_h - \widehat{\Delta}$, \widehat{C}_h est la variance empirique dans la tranche I_h , et $\widehat{\Delta} = \sum_{h=1}^H \frac{n_h}{n} \widehat{C}_h$.

$\widehat{\Sigma}$ est lui estimé par l'estimateur empirique usuel. L'estimateur \widehat{W} étant convergent à la vitesse \sqrt{n} , le sous-espace propre de $\widehat{\Sigma}^{-1}\widehat{W}$ de dimension K est lui-même un estimateur convergent de l'espace E.D.R. Cependant, les valeurs propres de $\widehat{\Sigma}^{-1}\widehat{W}$ sont presque sûrement distinctes. La méthode fournit donc une liste d'estimateurs de l'espace propre sans apporter d'information sur la dimension K qui joue pourtant un rôle crucial.

Remarque : Il est possible de fournir un cadre général pour les méthodes SIR I, SIR II et PHD, car dans ces trois cas, l'estimation des paramètres repose sur la

diagonalisation d'une matrice $\Sigma^{-1}W$, et d'une méthode à l'autre, seule W change. Par exemple, pour SIR I, W est une approximation par tranchage de $\text{var}(E(x/y))$ et, pour PHD, de $W = E(yx^t x)$.

2.2. Détermination de K pour SIR II

Pour $i = 1, \dots, K$, soient b_i le vecteur propre Σ -orthonormé et $P_i = b_i b_i^t \Sigma$ projecteur propre Σ -orthogonal associés à la valeur propre λ_i de $\Sigma^{-1}W$. Notons P_{K+1} le projecteur sur l'orthogonal de S_K , S_K étant l'espace engendré par b_1, \dots, b_K . Rappelons aussi que l'on suppose que $\lambda_{K+1} = \lambda_{K+2} = \dots = \lambda_p$. Soit $(\hat{\lambda}_i)_{i=1, \dots, p}$ la suite décroissante des valeurs propres de $\hat{\Sigma}^{-1}\hat{W}$ de vecteurs propres associés $\hat{\Sigma}$ -orthonormés $(\hat{b}_i)_{i=1, \dots, p}$. On rappelle que $\hat{\lambda}_i$ et \hat{b}_i convergent vers λ_i et b_i respectivement, pour $i = 1, \dots, K$ et $\hat{\lambda}_{K+1} = \frac{1}{p-K} \sum_{l=K+1}^p \hat{\lambda}_l$, et $\hat{P}_{K+1} = \frac{1}{p-K} \sum_{l=K+1}^p \hat{b}_l \hat{b}_l^t \hat{\Sigma}$ convergent respectivement vers λ_{K+1} et P_{K+1} .

Afin d'apprécier la qualité de l'estimation des sous-espaces propres de $\Sigma^{-1}W$, un critère a été introduit pour les méthodes SIR I et PHD par Ferré (1998). Ce dernier conduit à un choix de la dimension K à partir de la quantité :

$$R(q) = 1 - \frac{1}{q} E(\text{tr}(\hat{\Pi}_q \Pi_q)), q = 1, \dots, p,$$

où Π_q est le projecteur Σ -orthogonal sur le sous-espace engendré par b_1, \dots, b_q et $\hat{\Pi}_q$ son estimateur. Cette fonction de perte permet de mesurer les écarts respectifs entre les sous-espaces de l'espace E.D.R et leurs estimateurs.

Mais, pour des raisons pratiques, on ne peut utiliser le critère directement sous cette forme, puisque Π_q est a priori inconnu. On a alors recours à un développement limité de l'expression $E(\text{tr}(\hat{\Pi}_q \Pi_q))$ pour obtenir le terme principal dans le cas de SIR I (voir Ferré 1998) :

$$R(q) = \frac{1}{nq} \left[\sum_{i=1}^q \sum_{j=q+1}^K \frac{1}{(\lambda_j - \lambda_i)^2} (\lambda_i + \lambda_j + (-1 + \kappa)\lambda_i \lambda_j) + (p - K) \sum_{i=1}^q 1/\lambda_i \right]$$

avec, rappelons le, pour $i = 1, \dots, K$, λ_i est valeur propre de la matrice $\Sigma^{-1}W$ et κ est le paramètre de Kurtosis de la loi de x .

Dans le cas de SIR II, nous proposons ci-dessous le critère correspondant, obtenu en adaptant le critère $R(q)$. En particulier, l'intégration de la valeur propre multiple non nulle et l'utilisation de moments d'ordre supérieur à deux ne permettent pas l'obtention d'une expression aussi simple que dans le cas de SIR I.

Ainsi, sous les hypothèses suivantes :

1) pour n tel que, pour $i = 1, \dots, K + 1$:

$$\sqrt{\frac{\ln \ln(n)}{n}} < \frac{\inf_{i < j} (\lambda_i - \lambda_j)}{2C}$$

où C est une constante strictement positive telle que :

$$\left\| \sqrt{\frac{n}{\ln \ln(n)}} \left(\widehat{\Sigma}^{-1} \widehat{W} - \Sigma^{-1} W \right) \right\| < C \text{ p.s.}$$

2) x suit une loi elliptique symétrique;

3) conditionnellement à y , x suit une loi symétrique;

4) les moments d'ordre 4 de x , conditionnels ou non, existent;

on obtient le résultat suivant, en notant \otimes le produit de Kronecker de matrice et K_{p^2} la matrice de commutation (voir Magnus et Nedecker, 1987) :

Théorème 1. Pour $q = 1$ à $K - 1$:

$$\begin{aligned} R(q) &= \frac{1}{nq} \sum_{i=1}^q \sum_{j=q+1}^K \frac{1}{(\lambda_j - \lambda_i)^2} [(1 + \kappa)(\lambda_j + \lambda_i + \lambda_j \lambda_i) + A_{ij}] \\ &+ \frac{1}{nq} \sum_{i=1}^q \frac{1}{(\lambda_{K+1} - \lambda_i)^2} [(1 + \kappa)(p - K)(\lambda_{K+1} + \lambda_i + \lambda_{K+1} \lambda_i) + A_{iK+1}] \\ &+ 0(n^{-3/2}) \end{aligned}$$

$$\begin{aligned} \text{et } R(K) &= \frac{1}{nq} \sum_{i=1}^K \frac{1}{(\lambda_{K+1} - \lambda_i)^2} \\ &[(1 + \kappa)(p - K)(\lambda_{K+1} + \lambda_i + \lambda_{K+1} \lambda_i) + A_{iK+1}] + 0(n^{-3/2}), \end{aligned}$$

avec, pour $i = 1, \dots, q$ et $j = 1, \dots, K + 1$,

$$\begin{aligned} A_{ij} &= \text{tr}((P_j \Sigma^{-1} \otimes P_i \Sigma^{-1}) \\ &((I_{p^2} + K_{p^2})(1 + \kappa)((2 + \lambda_j + \lambda_i)G_H + G_H^2(\Sigma^{-1} \otimes \Sigma^{-1})) \\ &+ 4\kappa K_{p^2} G_H - A_H)) \end{aligned}$$

$$\text{où } A_H = \sum_{h=1}^H \frac{n_h}{n} \text{var}((K_h \Sigma^{-1} \otimes I_p + K_h \Sigma^{-1} \otimes I_p)E(x \otimes x/y)),$$

$$\text{et } G_H = \sum_{h=1}^H K_h \otimes K_h,$$

K_h ayant été défini au paragraphe précédent.

Des éléments de preuves sont donnés en Appendice.

Remarques

Tous les termes figurant dans ce théorème sont estimables en remplaçant chaque paramètre par sa version empirique usuelle. Ces estimateurs convergeant tous à la vitesse \sqrt{n} , l'estimateur de $R(q)$ est lui-même convergent à la vitesse \sqrt{n} .

On peut montrer aisément que l'estimateur obtenu est sans biais à l'ordre $O(n^{-3/2})$ et sa variance est un $O(n^{-3})$.

En pratique, $R(q)$ n'est pas directement utilisable, car K reste inconnu. Cependant, nous suggérons :

– soit de déterminer une valeur K^* , par exemple, par un test statistique (Schott(1994) propose un test dont la statistique suit asymptotiquement une loi de χ^2) et de remplacer K par K^* dans l'expression de $R(q)$, ce qui revient à corriger un phénomène de surparamétrisation dû au test.

– soit de faire varier K de 1 à p (en prenant q inférieur ou égale à K) et de travailler, comme avec SIR I avec un critère $R(q, K)$. C'est cette dernière démarche que nous retenons et nous déterminons la valeur de K à partir des p représentations des «courbes» $R(q, K)$. Comme nous le verrons dans les applications ci-après, les similitudes entre ces p courbes rendent possible une telle démarche.

On notera le rôle important joué par l'inverse des différences des valeurs propres successives. Cela est particulièrement intéressant ici, de part l'existence de la valeur propre multiple non nulle qui conduit à des valeurs propres observées, de tailles voisines, et donc à des valeurs élevées du critère.

Nous avons supposé les valeurs propres $(\lambda_1, \dots, \lambda_K)$ distinctes pour simplifier. Il est cependant possible d'obtenir une version de $R(q)$ dans le cas de valeurs propres multiples, mais son utilisation paraît délicate en raison de la remarque ci-dessus.

La démarche utilisée ici revient à évaluer la qualité d'estimation de chaque «sous-modèle». Il serait également possible de comparer chaque sous modèle estimé au «vrai modèle». Différentes mesures allant dans ce sens ont été proposées dans Ferré (1998).

3. Applications

3.1. Simulations

Pour les simulations nous avons utilisé le modèle suivant :

$$y = \exp(\beta_1^t x) - \exp(\beta_2^t x) + \exp(|\beta_3^t x|) + \beta_4^t x + \epsilon,$$

où $x \sim \mathcal{N}(0, I_7)$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$, $\beta_1 = (1, 0, 0, 1, 1, 0, 0)^t$, $\beta_2 = (0, 1, 1, 0, 0, 0, 0)^t$, $\beta_3 = (0, 0, 0, 0, 0, 1, -1)^t$, $\beta_4 = (0, 0, 0, 0, 0, 0, 1)^t$, et donc ici $K = 4$, $p = 7$. Le but de ces simulations est de s'assurer du bon fonctionnement du critère $R(q)$ et de comparer les résultats avec ceux de la SIR I. Nous avons donc simulé n valeurs du couple (x, y) au moyen d'un générateur de nombres pseudo-aléatoires et du modèle ci-dessus, avec $\sigma = 0.01$. Deux tailles d'échantillons ont été considérées : $n = 100$ et $n = 2000$. Dans chacun des cas, les méthodes SIR I et SIR II ont été mises en œuvre.

Dans un premier temps, pour $n = 100$, nous avons pris $H = 10$, où H désigne rappelons-le, le nombre de tranches du support de y utilisé pour estimer W (cf § 2.1). Les représentations des $R(q, K)$ sont données figure 1, la valeur de K étant portée sur la courbe et les valeurs de $R(q, K)$, $K < q$ étant arbitrairement posées égales à zéro. Les valeurs observées dans le cas de SIR I conduisent raisonnablement à retenir une seule dimension (quelle que soit la valeur de K , on a $R(1, K) \simeq 0.2$, alors que $R(2, K) \simeq 0.4$). Par contre, pour la SIR II, même si, à priori, une seule dimension est une solution convenable ($R(1, K) < 0.08$), au vu de la faible taille de l'échantillon, il semble possible d'étendre le choix à deux dimensions ($R(2, K) < 0.25$, mais $R(3, 5) < 0$). Le tableau 1, représentant le nuage de point des $\hat{\beta}^t x$ en fonction des $\beta^t x$, donne une idée de la qualité d'estimation de chacune des directions. Ainsi, pour la SIR I, seule la première dimension semble être convenablement estimée, alors que pour la SIR II, les deux premières semblent être satisfaisantes.

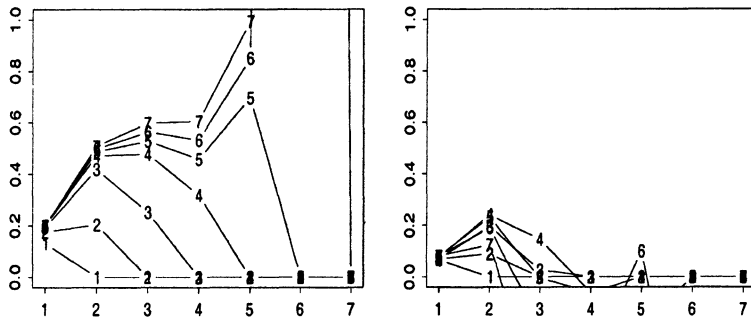


FIGURE 1

Représentation des 7 courbes $R(q, K)$, $K = 1, \dots, 7$, en fonction de $q = 1, \dots, K$, pour $n = 100$ et $H = 10$.

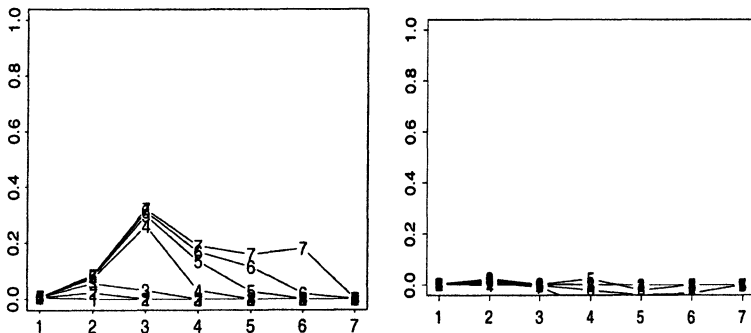


FIGURE 2

Représentation des 7 courbes $R(q, K)$, $K = 1, \dots, 7$, en fonction de $q = 1, \dots, K$, pour $n = 2000$ et $H = 50$.

Ensuite, nous avons considéré $n = 2\,000$ individus et $H = 50$ classes. Avec une telle taille d'échantillon, on peut s'attendre à une amélioration de la qualité de l'estimation des β_i . Ceci se confirme tableau 2, qui est l'équivalent du Tableau 1 pour notre nouvelle configuration. En effet, pour la SIR I, les deux premières directions sont mieux estimées que dans le cas précédent, sans toutefois atteindre la qualité obtenue pour les trois premières directions de SIR II. Ces résultats sont corroborés par les graphiques des $R(q, K)$ de la figure 2. Ceux-ci conduisent à choisir $K = 2$ dans le cas de SIR I et $K = 3$ dans celui de la SIR II. On notera que la quatrième dimension ne peut être estimée convenablement en raison de la faible variance de cette composante par rapport aux trois autres.

TABLEAU 1
Représentation des $\hat{\beta}_i^t x$ en

fonction des $\beta_i^t x$: $n = 100, H = 10$.

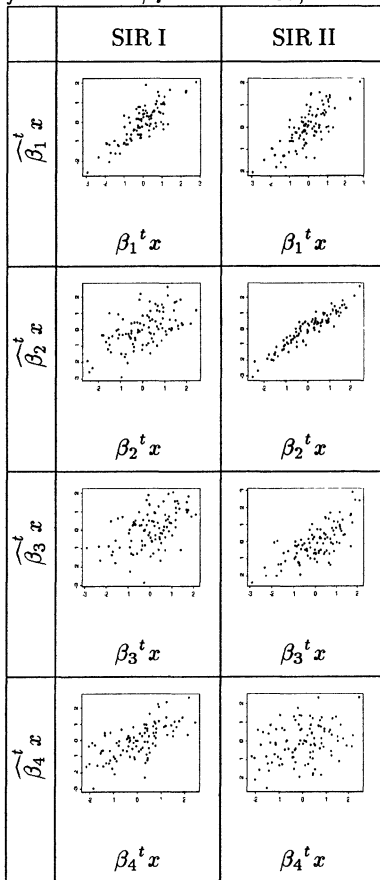
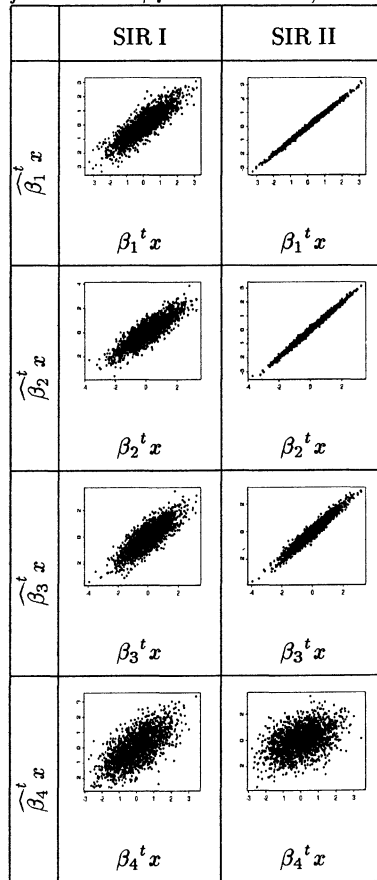


TABLEAU 2
Représentation des $\hat{\beta}_i^t x$ en

fonction des $\beta_i^t x$: $n = 2\,000, H = 50$.



3.2. Données réelles

Nous avons appliqué les méthodes SIR I et SIR II sur les données «Loves Data» présentées dans Zuroff (1994). Le but est d'étudier la relation entre certains caractères dépressifs et l'intensité d'une relation amoureuse. Les variables sont mesurées au travers d'échelles élaborées à partir de questionnaires. Ainsi, la variable y mesurant l'intensité amoureuse est un score global établi à partir du premier axe principal d'une A.C.P effectuée sur des échelles d'intensité amoureuse. Pour expliquer y , nous avons sélectionné dans le fichier initial sept variables de type psychologique mesurant :

x_1 : l'anxiété, x_2 : la peur d'être trop intime, x_3 : la névrose, x_4 : l'extraversion, x_5 : la capacité d'adaptation, x_6 : le charme, x_7 : la minutie.

Ces variables sont regroupées pour former la variable x . Les variables x_3 à x_7 correspondent à une auto-évaluation. L'échantillon de taille 159 a été pris parmi des étudiants de l'Université de Mc Gill au Canada.

Nous avons mis en œuvre les méthodes SIR I et SIR II, chacune en considérant dix classes. Les résultats obtenus montrent que, si la SIR I est inefficace, la SIR II, elle, conduit, à retenir trois dimensions (voir Figure 3). Les «mauvaises» valeurs obtenues pour le critère $R(q, K)$ en SIR I s'expliquent par la proximité des valeurs propres (tableau 3).

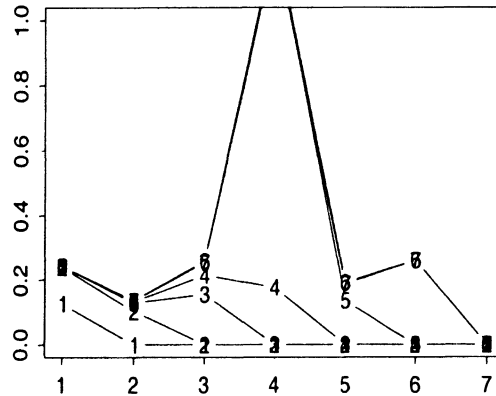


FIGURE 3

Représentation des 7 courbes $R(q, K)$, $K = 1, \dots, 7$, en fonction de $q = 1, \dots, 7$, pour les données réelles

TABLEAU 3

Estimation des valeurs propres pour les données réelles

SIR I	0,174	0,097	0,087	0,055	0,0452	0,028	0,005
SIR II	1,006	0,706	0,389	0,269	0,241	0,152	0,115

Vu le faible nombre d'individus, il ne paraît pas réaliste de mettre en œuvre l'estimation non-paramétrique de la fonction f avec $K = 3$. Cependant, il est possible d'exploiter ces résultats sur un plan descriptif. L'observation des coordonnées des vecteurs β , données Tableau 4, montre que, par exemple, «l'anxiété» et «la peur d'être trop intime» expliquent bien l'intensité affective de la relation; à l'inverse, «la capacité d'adaptation» ne semble pas intervenir. De plus, il est possible de représenter le nuage de points dans les directions qui expliquent au mieux la variable «intensité affective», pour interpréter le comportement des individus. Ces graphiques sont donnés dans le tableau 5. Par exemple, relevons que le nuage $(y, \beta_1^t x)$ met en évidence six individus ayant une intensité affective faible, un score d'anxiété plutôt fort et semblant craindre l'intimité. D'autre part, $(y, \beta_3^t x)$ révèle quatre individus névrosés, peu enclins à l'extraversion et de score d'intensité affective plutôt faible. Ici, la réduction de la dimension est intéressante car elle permet de passer du nuage de points initial en dimension 8 à un nuage de points en dimension 4.

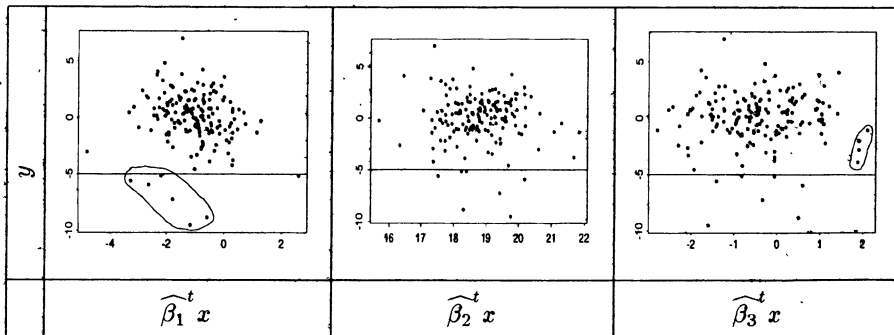
TABLEAU 4

Estimation des vecteurs propres associés aux trois plus grandes valeurs propres.

$\widehat{\beta}_1$	-0,101	0,118	-0,013	-0,034	0,015	0,027	0,030
$\widehat{\beta}_2$	-0,056	0,084	0,028	0,026	0,021	0,059	0,046
$\widehat{\beta}_3$	-0,036	-0,070	0,034	-0,031	0,008	0,017	-0,019

TABLEAU 5

Représentation de y en fonction de $\widehat{\beta}_i^t x$ pour les données réelles



Références

- [1] BRIEMAN L. & FRIEDMAN J. (1985), «Estimating Optimal Transformations for Multiple Regression and Correlation», *Journal of the American Statistical Association*, **80**, 580-597.
- [2] COOK R.D. & WEISBERG S. (1991), Comment on «Sliced Inverse Regression for Dimension Reduction», K.C. Li, *Journal of the American Statistical Association*, **86**, 328-332.
- [3] FERRÉ L. (1997), «Dimension Choice for Sliced Inverse Regression Based on Ranks», *Student*, **2**, 95-108.
- [4] FERRÉ L. (1998), «Determining the Dimension in Sliced Inverse Regression and Related Methods», *Journal of the American Statistical Association*, **93**, 132-140.
- [5] FRIEDMAN J. & Stuetzle W. (1981), «Projection Pursuit Regression», *Journal of the American Statistical Association*, **76**, 817-823.
- [6] HALL P. & LI K.C. (1993), «On Almost Linearity of Low Dimensional Projection from High Dimensional Data», *The Annals of Statistics*, **21**, 867-889.
- [7] HÄRDLE W. & STOKER T.M. (1989), «Investigating Smooth Multiple Regression by the Method of Average Derivatives», *Journal of the American Statistical Association*, **84**, 986-995.
- [8] HASTIE T. & TIBSHIRANI R. (1986), «General Additive Model», *Statistical Science*, **1**, 297-318.
- [9] HSING T. & CARROLL R.J. (1992), «An Asymptotic Theory for Sliced Inverse Regression», *The Annals of Statistics*, **20**, **2**, 1040-1061.
- [10] KÖTTER T. (1996), «An Asymptotic Result for Sliced Inverse Regression», *Computational Statistics*, **11**, 113-136.
- [11] LI K.C. (1991), «Sliced Inverse Regression for Dimension Reduction» (with comments), *Journal of the American Statistical Association*, **86**, 316-327.
- [12] LI K.C. (1992), «On Principal Hessian Directions for Data Visualisation and Dimension Reduction : another application of Stein's lemma», *Journal of the American Statistical Association*, **87**, 1025-1039.
- [13] MAGNUS J.R. & NEUDECKER H. (1987), *Matrix differential calculus with applications in statistics and econometrics*, J. Wiley, 1987.
- [14] SCHOTT J.R. (1994), «Determining the Dimensionality in Sliced Inverse Regression», *Journal of the American Statistical Association*, **89**, 141-148.
- [15] ZUROFF D.C. (1994), «Depressive Personality Styles and Adult Attachment Style», *Journal of Personality Assessment*, **63**, 453-472.

Appendice : Quelques éléments de démonstration

Les résultats s'obtiennent en appliquant le lemme 1 ci-dessous à la matrice $\widehat{V} = \widehat{W}$.

Lemme 1 (Ferré (1998)) Soit V et Σ deux matrices estimées par \widehat{V} et $\widehat{\Sigma}$ et soient $U_1 = \sqrt{n}(\widehat{V} - V)$ et $U_2 = \sqrt{n}(\widehat{\Sigma} - \Sigma)$. Sous les conditions (1), (2), (3), on a :

$$R(q) = \frac{1}{nq} \sum_{i=1}^q \sum_{j=q+1}^{K+1} \frac{1}{(\lambda_j - \lambda_i)^2} \text{tr}[(P_j \Sigma^{-1} \otimes P_i \Sigma^{-1})(\text{var}(U_1) + \lambda_j \lambda_i \text{var}(U_2) - (\lambda_i + \lambda_j) \text{cov}(U_1, U_2))] + 0(n^{-3/2}).$$

On calcule ensuite $\text{var}(U_1)$ à partir de l'expression de \widehat{W} et on obtient :

$$\text{var}(U_1) = \sum_{h=1}^H \frac{n_h}{n} [K_h \Sigma^{-1} \otimes I_p + K_h \Sigma^{-1} \otimes I_p] E(\text{var}(xx^t/y)) [K_h \Sigma^{-1} \otimes I_p + K_h \Sigma^{-1} \otimes I_p] + G_H \text{var}(U_2) G_H^t.$$

Par ailleurs, comme x suit une loi elliptique symétrique, on a :

$$\text{var}(U_2) = (1 + \kappa)(I_{p^2} + K_{p^2})(\Sigma \otimes \Sigma) + \kappa \text{vec}(\Sigma) \text{vec}(\Sigma)^t.$$

Un calcul rapide conduit à :

$$\text{cov}(U_1, U_2) = -G_H \text{var}(U_2).$$

En remplaçant les variances et covariances dans les différents termes de $R(q)$, on obtient : pour $j = 1, \dots, K$

$$\begin{aligned} *tr((P_j \Sigma^{-1} \otimes P_i \Sigma^{-1}) \text{var}(U_1)) &= (1 + \kappa)[(\lambda_i + \lambda_j) + \\ &+ 2tr((I_{p^2} + K_{p^2})(P_j \Sigma^{-1} \otimes P_i \Sigma^{-1}) G_H)] \\ &+ 4\kappa tr(K_{p^2}(P_j \Sigma^{-1} \otimes P_i \Sigma^{-1}) G_H) \\ &tr((P_j \Sigma^{-1} \otimes P_i \Sigma^{-1})(-A_H + \\ &(1 + \kappa)(I_{p^2} + K_{p^2})(G_H^2(\Sigma^{-1} \otimes \Sigma^{-1}))), \end{aligned}$$

$$*tr((P_j \Sigma^{-1} \otimes P_i \Sigma^{-1}) \text{var}(U_2)) = (1 + \kappa),$$

$$\begin{aligned} *tr((P_j \Sigma^{-1} \otimes P_i \Sigma^{-1}) \text{cov}(U_1, U_2)) &= -(1 + \kappa) \\ &(tr((I_{p^2} + K_{p^2})(P_j \Sigma^{-1} \otimes P_i \Sigma^{-1}) G_H), \end{aligned}$$

et, pour $j = K + 1$,

$$\begin{aligned} *tr((P_{K+1}\Sigma^{-1} \otimes P_i\Sigma^{-1})\text{var}(U_1)) &= (1 + \kappa)[(p - K)(\lambda_i + \lambda_{K+1}) + 2tr(\\ &\quad (I_{p^2} + K_{p^2})(P_{K+1}\Sigma^{-1} \otimes P_i\Sigma^{-1})G_H)] \\ &\quad + 4\kappa tr(K_{p^2}(P_{K+1}\Sigma^{-1} \otimes P_i\Sigma^{-1})G_H) \\ &\quad - tr((P_{K+1}\Sigma^{-1} \otimes P_i\Sigma^{-1})(A_H + \\ &\quad (1 + \kappa)(I_{p^2} + K_{p^2})(G_H^2(\Sigma^{-1} \otimes \Sigma^{-1}))))), \end{aligned}$$

$$*tr((P_{K+1}\Sigma^{-1} \otimes P_i\Sigma^{-1})\text{var}(U_2)) = (1 + \kappa)(p - K),$$

$$*tr((P_{K+1}\Sigma^{-1} \otimes P_i\Sigma^{-1})\text{cov}(U_1, U_2)) = -(1 + \kappa)tr(\\ (I_{p^2} + K_{p^2})(P_{K+1}\Sigma^{-1} \otimes P_i\Sigma^{-1})G_H).$$

Le résultat s'obtient en intégrant ces expressions dans $R(q)$ et en arrangeant.