

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.

Cependant, pour beaucoup d'applications, la mise en évidence d'un modèle paramétrique adéquat n'est pas simple. Aussi, quand un modèle paramétrique n'est pas disponible, les techniques de régression nonparamétriques apparaissent comme une alternative qui offre la flexibilité désirée dans la modélisation :

$$y = f(x) + \epsilon.$$

Le thème commun de la régression fonctionnelle est l'idée d'un lissage local qui n'exploite que le postulat de continuité ou de dérivabilité de la fonction inconnue de régression f . La qualité du lissage local en un point dépend alors de la présence de suffisamment de données dans le voisinage de ce point. Lorsque la variable explicative est unidimensionnelle ($p = 1$), nous pouvons citer parmi tant d'autres, la méthode des noyaux, les splines de lissage (voir par exemple Eubank (1988) ou Haerdle (1990)). Mais lorsque la dimension p de x devient grande, les méthodes nonparamétriques nécessitent des échantillons de grande taille pour l'estimation de f , voir Haerdle (1990).

Pour surmonter ce problème de dimension, des modèles de régression semiparamétrique ont été développés. Nous nous intéressons ici au modèle semiparamétrique de régression proposé par Li (1991) :

$$y = f(x'\beta_1, \dots, x'\beta_K, \epsilon)$$

où x est un vecteur aléatoire, l'erreur ϵ est indépendante de x et de loi inconnue et arbitraire, la *fonction de lien* f est inconnue et arbitraire, et K est strictement inférieur à la dimension de x . Nous considérons en particulier l'estimation de la partie paramétrique de ce modèle. Duan et Li (1991) et Li (1991) ont présenté une méthode permettant d'estimer une base de l'espace engendré par β_1, \dots, β_K . Ils ont appelé cette méthode "Sliced Inverse Regression" (SIR), que l'on peut traduire en français par "régression inverse par tranches". Cette méthode ne nécessite aucune spécification en ce qui concerne la distribution de l'erreur et la fonction de lien. Elle repose sur une propriété de la fonction de régression inverse. Ainsi, au lieu de régresser y sur x , on régresse x sur y . Un premier avantage de cette inversion de rôle est que l'on a réduit la dimension du problème: nous avons en effet maintenant p problèmes de dimension 1, la régression inverse permettant de régresser chaque coordonnée de x sur y . Après avoir obtenu une estimation de la partie paramétrique, on peut utiliser les techniques usuelles de lissage pour estimer la fonction de lien f , techniques d'autant plus efficaces que K est petit.

Avant de présenter plus en détail la méthode SIR, de brèves réflexions d'une part sur le nom de la méthode, d'autre part sur le principe d'estimation qu'elle met en oeuvre nous paraissent nécessaires.

Le nom de la méthode, SIR, fait référence à deux aspects : (1) une propriété géométrique essentielle de la courbe de régression inverse faisant intervenir divers moments du couple (y, x') ; (2) une discrétisation (slicing) de la variable dépendante qui ne modifie pas la partie paramétrique du modèle mais qui simplifie considérablement l'estimation des moments intervenant dans la propriété géométrique. Mais nous verrons d'une part que l'on peut combiner plusieurs discrétisations et d'autre part que l'on peut estimer directement les moments. Toutes les méthodes d'estimation que nous envisageons sont des méthodes de substitution de moments empiriques aux moments théoriques.

Dans la section 2, nous donnons dans un premier temps une description détaillée des modèles de régression semiparamétrique de référence (cas univarié et cas multivarié). Nous décrivons ensuite dans les sections 3 et 4 les méthodes SIR disponibles dans ces deux cadres.

2 Modèles semiparamétriques de référence

Nous décrivons les modèle de référence et précisons les paramètres d'intérêt que l'on est amené à estimer. Nous donnons ensuite une condition fondamentale sur la distribution de la variable explicative, condition indispensable dans la théorie de SIR, et nous discutons les éventuelles restrictions pratiques que pourrait entraîner une telle hypothèse.

2.1 Cas où la variable y est unidimensionnelle

Le modèle de référence sur lequel on travaille est le modèle de régression semiparamétrique suivant qui a été proposé par Li (1991):

$$y = f(x'\beta_1, \dots, x'\beta_K, \epsilon) \quad (M1)$$

où:

- y est une variable unidimensionnelle à expliquer;
- x est une variable aléatoire explicative de dimension p ayant une espérance μ , une matrice de covariances Σ définie positive et vérifiant une condition fondamentale précisée ci-après;
- β_1, \dots, β_K sont des vecteurs de paramètres réels de dimension p , inconnus et linéairement indépendants, avec $K < p$;
- ϵ est un terme d'erreur aléatoire de distribution inconnue et ϵ est indépendante de x ;
- f est le paramètre fonctionnel à valeur dans \mathcal{R} , inconnu et arbitraire.

K étant supposé strictement inférieur à p , ce modèle permet une réduction de dimension de l'espace des variables explicatives. Dans un tel modèle, on appelle *indices* les termes $x'\beta_k$, $k = 1, \dots, K$. Ainsi, pour $K > 1$, on parle de modèle multi-indices et lorsque $K = 1$, de modèle à un seul indice $x'\beta$.

De nombreux modèles de régression de la littérature statistique sont des cas particuliers du modèle (M1), notamment,

- le modèle de régression linéaire ($K = 1$): $f(x'\beta, \epsilon) \equiv \alpha + x'\beta + \epsilon$;
- le modèle avec erreur additive ($K = 1$): $f(x'\beta, \epsilon) \equiv h(x'\beta) + \epsilon$, où la fonction h est inconnue;

- le modèle avec erreur multiplicative ($K = 1$): $f(x'\beta, \epsilon) \equiv \alpha + \epsilon h(x'\beta)$, avec la fonction h inconnue;
- "Projection Pursuit Regression": $f(x'\beta_1, \dots, x'\beta_K, \epsilon) \equiv h_1(x'\beta_1) + \dots + h_K(x'\beta_K) + \epsilon$, avec les fonctions h_j inconnues, voir Hall (1989) pour une revue bibliographique sur le sujet;
- le modèle avec hétérogénéité ($K = 2$): $f(x'\beta_1, x'\beta_2, \epsilon) \equiv h_1(x'\beta_1) + \epsilon h_2(x'\beta_2)$, les fonctions h_1 et h_2 étant inconnues.

D'autres modèles rentrent dans le cadre du modèle (M1). On peut citer par exemple les modèles à choix discret comme le modèle Probit ou le modèle Logit, ou encore les modèles de sélection tel que le modèle Tobit : pour une présentation détaillée de ces modèles, voir Gouriéroux (1989).

En permettant à la fonction f et à l'erreur ϵ d'être complètement non spécifiées, on n'impose aucune hypothèse structurelle sur la forme du lien qui existe entre y et les K régresseurs réduits $x'\beta_1, \dots, x'\beta_K$. Ainsi dans le cadre du modèle de référence (M1), l'étude de la relation entre y et x se réduit à l'examen de la structure qui existe entre y et ses K régresseurs $x'\beta_1, \dots, x'\beta_K$.

Cependant, il convient de remarquer que sous les hypothèses du modèle (M1), les paramètres β_k , $k = 1, \dots, K$, ne sont pas totalement identifiables. En effet, soit B la matrice de dimension $p \times K$: $[\beta_1, \dots, \beta_K]$. Pour toute matrice A inversible d'ordre K , on peut réécrire (M1) sous la forme :

$$y = g(B'x, \epsilon) = g\left((A')^{-1}(BA)'x, \epsilon\right).$$

Etant donné que la fonction f est arbitraire, le terme $(A')^{-1}$ peut être "absorbé" par f . Ainsi, le modèle (M1) ne permet pas de distinguer les vecteurs colonnes de B de ceux de la matrice BA . Il s'ensuit que l'on ne peut donc au mieux qu'estimer le sous-espace engendré par les colonnes de B .

Définition. Notons $E = \text{Vect}(B)$ le sous-espace linéaire de \mathbb{R}^p de dimension $K < p$ engendré par les colonnes de la matrice B , c'est-à-dire par les β_k , $k = 1, \dots, K$.

On appelle E l'espace EDR ("effective dimension reduction"). De plus, toute combinaison linéaire des β_k appartient à E et est appelée *direction EDR*.

L'objectif majeur que l'on se fixe dans ce modèle est l'estimation de la partie paramétrique, c'est-à-dire l'estimation d'une base de l'espace EDR E . L'estimation de la partie fonctionnelle f peut être obtenue avec les méthodes nonparamétriques usuelles (méthodes des noyaux, des splines de lissage ou de régressions ou des ondelettes).

Quelques remarques sur le modèle.

- (i) Quand $K = 1$, on s'intéresse simplement à l'estimation de la direction du vecteur β . Duan et Li (1991) ont étudié un tel modèle, dit à un seul indice.
- (ii) Pour que les β_k soient individuellement identifiables, il est nécessaire d'imposer en plus des contraintes de normalisation, d'orthogonalité et de sens aux β_k .

(iii) Cette limitation concernant la non identifiabilité des β_k n'affecte cependant pas la spécification du modèle dans l'étape suivant la réduction de dimension. En effet, toute base de l'espace EDR E peut être ensuite utilisée pour l'étape d'estimation du paramètre fonctionnel f .

(iv) Dans le modèle de référence (M1), on suppose K , le paramètre de réduction de dimension, connu. Divers travaux ont été effectués sur le choix de la dimension K quand on est confronté à un échantillon, voir section 3.1.4.

Une condition fondamentale et discussion Le théorème de caractérisation de l'espace EDR (base des méthodes SIR) énoncé à la partie suivante nécessite que x vérifie la condition suivante: *la variable explicative x suit une distribution de probabilité non dégénérée telle que, pour tout $b \in \mathbb{R}^p$, l'espérance conditionnelle $E[b'x|\beta'_1x, \dots, \beta'_Kx]$ est linéaire en $\beta'_1x, \dots, \beta'_Kx$.*

Il est intéressant de noter que cette condition est vérifiée lorsque la distribution de x est de type elliptique. On peut donc la remplacer par l'hypothèse plus forte (mais plus simple à interpréter) suivante: *La variable explicative x suit une distribution elliptique.* On rappelle que les distributions multinormales, les distributions multidimensionnelles de Laplace, de Student et de Cauchy ainsi que la loi multinormale contaminée dans la variance, sont des exemples de distributions elliptiques.

Avant de poursuivre la présentation des méthodes SIR, il apparaît intéressant de discuter cette condition fondamentale sur laquelle est fondée la théorie de SIR. Remarquons tout d'abord qu'il est très difficile de vérifier a priori si la condition est satisfaite vu qu'elle dépend de directions inconnues. De plus, la plupart des données collectées ne suivent pas une distribution elliptique.

Lorsque la dimension de la variable explicative x n'est pas trop importante et que les données ne suivent pas une distribution elliptique, l'idée est d'essayer de forcer ces données à se comporter comme si tel était le cas. Si les données n'ont pas encore été collectées, il est possible par exemple, grâce à un plan d'expérience convenable, d'atteindre l'objectif recherché. Lorsque les données ont déjà été collectées, on dispose alors d'un échantillon de taille n . Diverses techniques ont été proposées pour forcer les données à être elliptiques. Citons par exemple la méthode du rééchantillonnage normal de Brillinger (1983) et une méthode plus récente due à Cook et Nachtsheim (1994).

Lorsque le nombre de variables explicatives est très grand, il devient très difficile de forcer les données à être elliptiques avec des méthodes du type de celles présentées précédemment. Cependant, Hall et Li (1993) ont montré par un argument Bayésien que la condition fondamentale est presque toujours approximativement vérifiée. Ils ont ensuite précisé comment ce résultat permet d'élargir le champ d'applicabilité de la régression inverse par tranches. Ainsi, pour de nombreux ensembles de données, une application "aveugle" des procédures de régression (SIR ou autre), sans vérifier la condition fondamentale, peut être utile et fournir encore une réponse à peu près correcte pour l'estimation des directions EDR. Cependant un diagnostic de vérification est recommandé après cette utilisation de SIR afin de regarder si, parmi les directions trouvées, certaines ne s'éloignent pas sérieusement de l'ellipticité. Et une fois que l'on a repéré de telles directions, on peut par exemple appliquer la méthode de Cook et Nachtsheim (1994) pour forcer l'ellipticité des données

sur ces directions-là.

Pour conclure cette discussion, il est intéressant de noter que dans le cas d'un modèle à un seul indice, Duan et Li (1991) ont étudié le comportement de l'estimateur $\hat{\beta}$ de la direction de β lorsque l'hypothèse d'ellipticité de la distribution de x était violée. Ils ont établi une borne pour le biais qui apparaît lorsque la distribution de x n'est pas elliptique. Enfin, on peut aussi préciser que Li (1991) présente des résultats obtenus par simulations sur des échantillons pour lesquels la variable explicative x simulée n'est pas elliptique (elle suit une distribution uniforme sur un hypercube). Il obtient cependant des estimations de bonne qualité.

2.2 Cas où la variable y est multidimensionnelle

Dans cette partie, on considère un modèle semiparamétrique dans lequel la variable dépendante et la variable explicative sont multidimensionnelles ($y \in \mathbb{R}^q$ et $x \in \mathbb{R}^p$) de la forme suivante :

$$y_j = f_j(x'\beta_1, \dots, x'\beta_K, \epsilon), \quad \text{pour } j = 1, \dots, q \quad (M3)$$

où :

- y_j est la $j^{\text{ème}}$ composante de la variable dépendante y de dimension q ;
- x est une variable aléatoire explicative de dimension p , d'espérance μ et de matrice de covariances Σ définie positive qui vérifie la condition fondamentale rappelée ci-après ;
- β_1, \dots, β_K sont des vecteurs de paramètres inconnus de dimension p , linéairement indépendants ;
- ϵ est une erreur aléatoire indépendante de x , de distribution inconnue ;
- les f_j , $j = 1, \dots, q$ sont les paramètres fonctionnels à valeur dans \mathbb{R} , et sont inconnus et arbitraires.

Tout comme dans le cas du modèle univarié et multi-indices, il est nécessaire de rajouter à ce modèle la condition fondamentale qui porte sur la distribution de la variable explicative x pour que la courbe de régression inverse possède des propriétés permettant la caractérisation de l'espace EDR E .

On s'intéresse toujours à la partie paramétrique de ce modèle. L'objectif des méthodes présentées SIR dans le cadre de ce modèle est l'estimation d'une base de l'espace EDR E . On se limite à l'estimation d'une base de E car, pour les mêmes raisons que celles décrites dans le cas d'un modèle univarié, les β_k ne sont pas totalement identifiables et on ne peut au mieux qu'estimer une base de l'espace engendré par les β_k , c'est-à-dire une base de E .

3 Méthodes SIR univariées (cas où $y \in \mathbb{R}$)

3.1 Philosophie et théorie de la méthode SIR

Nous étudions maintenant une propriété géométrique des courbes de régression inverse dans le cadre du modèle de référence (M1) et sous la condition fondamentale, propriété permettant de caractériser l'espace EDR E . Nous en déduisons ensuite une méthode d'estimation d'une base de l'espace EDR dans laquelle l'utilisation d'un découpage du support de y permet de simplifier grandement l'estimation des différents moments intervenant dans la partie théorique.

3.1.1 Régression inverse et modèle de référence

On considère la version standardisée z de x :

$$z = \Sigma^{-1/2}(x - \mu).$$

On peut alors réécrire le modèle (M1) sous la forme :

$$y = g(z'\eta_1, \dots, z'\eta_K, \epsilon) \quad (M2)$$

où $\eta_k = \Sigma^{1/2}\beta_k$, $k = 1, \dots, K$. (La fonction g est identique à la fonction f à une translation près de ses K premiers arguments.)

Notons N la matrice de dimension $p \times K$: $[\eta_1, \dots, \eta_K]$ et l'espace EDR standardisé $E^s = \text{Vect}(N)$. On appelle alors *direction EDR standardisée*, toute direction appartenant à cet espace.

Remarquons d'abord que si l'on considère une transformation monotone T de y , le modèle (M2) peut alors s'écrire :

$$T(y) = g^*(z'\eta_1, \dots, z'\eta_K, \epsilon) \quad (M2^*)$$

où g^* est la fonction composée de T et de g . Ainsi $T(y)$ obéit à un modèle de type (M2).

Définition. Étant donnée une telle transformation T , on appelle *courbe de régression inverse standardisée*, la fonction de \mathbb{R} dans \mathbb{R}^p définie par :

$$y \longrightarrow E[z|T(y)].$$

Le théorème ci-dessous établi par Li (1991) est le fondement de la méthode SIR.

Théorème de caractérisation de l'espace EDR standardisé E^s . Dans le cadre du modèle (M2) et sous la condition fondamentale, la courbe de régression inverse standardisée appartient à l'espace E^s , sous-espace de \mathbb{R}^p de dimension K engendré par les η_k .

Il s'en suit (voir Li (1991)) que la matrice de covariances de cette courbe,

$$\Gamma_T := \text{var}(E[z|T(y)]),$$

est dégénérée dans toute direction orthogonale aux η_k . Ainsi, les K vecteurs propres, v_1, \dots, v_K , associés aux K valeurs propres non nulles de la matrice Γ_T engendrent l'espace EDR standardisé E^s .

L'espace EDR E cherché est alors obtenu à partir de cet espace EDR standardisé: il est engendré par les K vecteurs $\Sigma^{-1/2}v_k$.

Remarque. Dans le théorème de caractérisation, la courbe de régression inverse centrée appartient à E^s mais elle n'engendre pas nécessairement ce sous-espace. Il peut ainsi parfois arriver que ces courbes n'engendrent que des sous-espaces de dimension strictement inférieure à K , sous-espaces ne contenant peut-être que peu d'information sur les paramètres. Ce point pathologique de la méthode est discuté par Li (1991) et Cook and Weisberg (1991) et des remèdes sont proposés: voir les méthodes SIR II et SAVE.

La monotonie de la transformation T nous assure que le modèle $(M1^*)$ n'est pas un modèle pathologique (au sens de Cook and Weisberg (1991)) quand le modèle initial $(M1)$ n'est pas lui-même pathologique.

3.1.2 Algorithme général des méthodes SIR

D'après les résultats théoriques de la section précédente, l'espace EDR standardisé E^s est indépendant du choix de la transformation monotone T , seule la base de cet espace est modifiée.

En pratique, on dispose d'un échantillon $\{(y_i, x_i'), i = 1, \dots, n\}$. On désire estimer une base de l'espace EDR E . On est alors amené à estimer la matrice Γ_T et ainsi à choisir une transformation T telle que cette matrice soit facilement estimable. Nous donnons ci-dessous l'algorithme des méthodes d'estimation des directions EDR. Dans la suite, différents choix de T seront discutés et en particulier, à la section suivante, le cas où T est un *découpage* du support de y , choix initial de la méthode SIR.

Etape I: Standardisation des x_i , $i = 1, \dots, n$.

$$z_i = \hat{\Sigma}^{-1/2}(x_i - \bar{x}), \quad i = 1, \dots, n,$$

où \bar{x} et $\hat{\Sigma}$ sont respectivement la moyenne empirique et la matrice de covariances empirique de l'échantillon des x_i :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

Etape II: Estimation $\hat{\Gamma}_T$ de Γ_T . Cette étape sera détaillée ultérieurement en fonction des différents choix de T .

Etape III: Décomposition spectrale de $\hat{\Gamma}_T$. Notons \hat{v}_j , $j = 1, \dots, p$ les p vecteurs propres de $\hat{\Gamma}_T$, classés par ordre décroissant de leurs valeurs propres respectives. Les K premiers sont des directions EDR standardisées estimées.

Etape IV : Construction des K directions EDR estimées.

$$\hat{b}_k = \hat{\Sigma}^{-\frac{1}{2}} \hat{v}_k, \quad k = 1, \dots, K.$$

3.1.3 Un choix particulier de T : le découpage en tranches

Initialement la régression inverse par tranches prend pour transformation T une discrétisation de y en H tranches distinctes :

$$s_1, \dots, s_h, \dots, s_H.$$

Le choix et la construction des tranches sont discutés ci-dessous.

Dans ce cadre, en notant $p_h = \text{Pr}(y \in s_h)$ et $m_h = E[z|y \in s_h]$, $h = 1, \dots, H$, la matrice Γ_T s'écrit très simplement sous la forme

$$\Gamma_T = \sum_{h=1}^H p_h m_h m_h'.$$

L'estimation des différentes grandeurs intervenant dans Γ_T ne pose aucune difficulté et correspond à l'étape II de l'algorithme précédent :

- p_h est estimée par la proportion empirique des y_i tombant dans la tranche s_h :

$$\hat{p}_h = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[y_i \in s_h], \quad h = 1, \dots, H.$$

- m_h est estimée par la moyenne empirique par des z_i tels que y_i appartiennent à la tranche s_h :

$$\hat{m}_h = \frac{1}{n \hat{p}_h} \sum_{i=1}^n z_i \mathbb{I}[y_i \in s_h], \quad h = 1, \dots, H.$$

- une estimation de la matrice Γ_T est alors donnée par :

$$\hat{\Gamma}_T = \sum_{h=1}^H \hat{p}_h \hat{m}_h \hat{m}_h'.$$

Choix du nombre H de tranches. Le nombre H peut être choisi arbitrairement. Les deux cas extrêmes qui consisteraient à prendre $H = 1$ ou $H = n$ (c'est-à-dire une seule tranche ou une seule observation par tranche) n'ont pas d'intérêt. En effet, dans le premier cas, la matrice $\hat{\Gamma}_T$ est égale à la matrice nulle et dans le second cas, $\hat{\Gamma}_T = I_p$. Entre ces deux valeurs particulières de H , il reste une grande liberté de choix pour H . Il convient cependant de noter que plus H augmente, plus le nombre de vecteurs propres non triviaux augmente. Ainsi, il apparaît nécessaire de choisir H suffisamment grand pour retrouver toutes les directions EDR. En particulier, pour ne pas faire une réduction de dimension artificielle, il est indispensable que le nombre de tranche H soit supérieur à K , la dimension de réduction du modèle. En pratique, d'après des observations faites sur des simulations, le choix de H n'a pas trop d'impact sur la qualité des estimations quand on

travaille avec des échantillons de taille $n \geq 100$. Un bon principe consiste à prendre, dans ce cas, un nombre de tranches suffisant sans être trop important en fonction de la taille de l'échantillon (par exemple 10 tranches pour $100 \leq n \leq 1000$ lorsque l'on prend $K < 10$), ce qui permet d'avoir suffisamment d'individus par tranches. Cependant, l'arbitraire qui peut présider au choix du nombre et de la position des tranches peut sensiblement dégrader la qualité des estimations lorsque l'on est en présence d'échantillons de petite taille ($n \leq 50$). Il est donc intéressant de considérer des solutions alternatives moins arbitraires. Ce thème est discuté à la section 3.2.

En fait, le choix de H est beaucoup moins crucial que le choix du paramètre de lissage en régression nonparamétrique. En théorie, un mauvais choix pour le paramètre de lissage en estimation fonctionnelle de la densité ou de la régression peut entraîner une vitesse de convergence très faible alors que pour la méthode SIR, quel que soit le choix de H , on a toujours une convergence en probabilité des directions EDR estimées vers de vraies directions EDR à la vitesse $1/\sqrt{n}$, voir Li (1991).

Choix des H tranches. Le problème du choix du nombre H de tranches ayant été discuté, la question qui se pose maintenant est de savoir comment choisir et construire ces H tranches. Deux options se présentent : soit prendre des tranches de même amplitude, soit faire des tranches contenant à peu près le même nombre d'individus.

Il apparaît préférable de prendre des tranches d'effectifs proches, ce qui permet de s'assurer d'avoir des tranches avec suffisamment d'individus. Une technique pour construire de telles tranches à partir d'un échantillon est la suivante. Notons \hat{r}_j , ($j = 1, \dots, H-1$) le (j/H) ème quantile empirique des y_i . Les tranches sont alors définies par :

$$s_1 = \{y_i | y_i \leq \hat{r}_1\}, \dots, s_h = \{y_i | \hat{r}_{h-1} < y_i \leq \hat{r}_h\}, \dots, s_H = \{y_i | y_i \geq \hat{r}_{H-1}\}.$$

Remarque. L'algorithme décrit précédemment est facile à mettre en oeuvre au niveau informatique. De plus, on peut remarquer que la méthode SIR ne nécessite aucun calcul itératif mais uniquement des calculs matriciels numériquement très rapides, indépendamment de la taille de l'échantillon.

3.1.4 Détermination de la dimension K du modèle

Jusqu'à présent, on a supposé que la dimension K de réduction du modèle était connue. Cependant dans la plupart des applications, cette dimension K est inconnue et doit être par conséquent estimée à partir de l'échantillon dont on dispose. Différents travaux permettant de déterminer la dimension K ont été publiés. La première approche de ce problème est fondée sur les tests d'hypothèses emboîtées : Li (1991) se place dans le cas où x est normalement distribuée et Schott (1994) dans le cas où la distribution de x est elliptique. La seconde approche proposée par Ferré (1996, 1998) repose sur la qualité de l'estimation des espaces de réduction effective de dimension évaluée par une mesure de l'écart entre l'espace EDR et son estimation. Il utilise alors cette mesure pour déterminer en pratique la dimension K à retenir.

3.1.5 Résultats asymptotiques

Nous ne donnons pas ici les différents résultats asymptotiques obtenus pour la méthode SIR "classique" (par opposition aux méthodes SIR alternatives présentées ci-après). Il nous semble cependant intéressant de noter que la vitesse de convergence des directions EDR estimées vers de vraies directions EDR est $n^{-1/2}$ (voir Li (1991)) et que cette vitesse est la même que celle obtenue dans le cadre d'un modèle paramétrique, alors que l'on travaille ici avec un modèle semiparamétrique. Pour plus de détails sur les résultats de convergence et de normalité asymptotique des directions EDR estimées, nous pouvons nous référer par exemple à Duan et Li (1991), Li (1991), Carroll et Li (1992), Saracco (1997).

3.2 Méthodes alternatives aux découpages

Nous avons mentionné, lors de la présentation de la méthode de Régression Inverse par Tranches, que pour des échantillons de taille $n \geq 100$, la méthode SIR fournit de très bonnes estimations des directions EDR, estimations qui ne sont pas sensibles au choix arbitraire des découpages fixés par l'utilisateur. Cependant, il arrive fréquemment que l'on ait à traiter en pratique des échantillons de faible taille (de l'ordre de 25 à 50 individus). On constate alors que la méthode SIR donne des résultats instables, cette instabilité provenant du choix des tranches. Ainsi, il apparaît intéressant de rechercher des méthodes alternatives à un découpage en tranches ("slicing") particulier fixé par l'utilisateur.

Toutes les méthodes alternatives que nous allons examiner se ramènent à la décomposition aux valeurs propres d'un estimateur de la matrice

$$\Gamma_T = \text{var}(E[z|T(y)])$$

pour différents choix de T .

La première méthode alternative, appelée Pooled Slicing, combine les résultats de plusieurs découpages différents du support de y . Pour les deux autres méthodes alternatives, la méthode de Hsing et Carroll (1992) et la méthode de régression inverse par lissage, la transformation T choisie est la fonction identité.

Nous ne précisons pas les résultats asymptotiques de chacune de ces méthodes, pour plus de détails à ce sujet, il faut se référer aux différents articles cités à chaque sous-section.

3.2.1 La méthode du Pooled-Slicing

Dans le but d'améliorer les estimations des directions EDR et d'éviter un choix de découpage particulier demandé à l'utilisateur, une solution est de combiner les résultats provenant d'un balayage de plusieurs découpages différents.

On considère ici D discrétisations T_d du support y . Pour chacune de ces discrétisations, on note $\Gamma_d = \text{cov}(E[z|T_d(y)])$. Soit $\{w_d\}_{d=1, \dots, D}$ des poids positifs tels que $\sum_{d=1}^D w_d = 1$. On définit alors la matrice agrégée

$$\Gamma^P = \sum_{d=1}^D w_d \Gamma_d.$$

Aragon et Saracco (1997) ont montré que les K vecteurs propres associés aux valeurs propres non nulles de Γ^P sont des directions EDR standardisées.

Pour l'estimation de Γ^P , on estime d'abord chaque matrice Γ_d de manière analogue à l'étape II de la méthode SIR classique pour chaque discrétisation T_d et après un choix des poids w_d , on obtient

$$\hat{\Gamma}^P = \sum_{d=1}^D w_d \hat{\Gamma}_d.$$

Choix des poids w_d . Le choix le plus simple est celui de l'équipondération: $w_d = 1/D$, $d = 1, \dots, D$. Un autre choix intéressant est de prendre une pondération proportionnelle à la somme des K plus grandes valeurs propres de chaque matrice $\hat{\Gamma}_d$.

Choix des D découpages. Pour balayer automatiquement un nombre important de découpages différents, il suffit d'imposer les trois contraintes suivantes: (i) un nombre minimum n_{\min} d'individus par tranche, (ii) un nombre minimum H_{\min} de tranches et (iii) construire des tranches de poids constant, c'est-à-dire contenant à peu près le même nombre d'individus pour chaque découpage. Ces trois contraintes permettent de déterminer les D découpages.

Par exemple, pour un échantillon de taille $n = 75$, si l'on choisit $n_{\min} = 5$ et $H_{\min} = 5$, la méthode du Pooled Slicing va balayer $D = 11$ découpages différents:

- $d = 1$: $H_{\min} = 5$ tranches de 15 individus,
- $d = 2$: 6 tranches dont 3 de 13 individus et 3 de 12 individus,
- $d = 3$: 7 tranches dont 5 de 11 individus et 2 de 10 individus,
- \vdots
- $d = 11$: 15 tranches de $n_{\min} = 5$ individus.

3.2.2 Méthode de Hsing et Carroll (1992)

Hsing et Carroll (1992) et Zhu et Ng (1995) se sont intéressés non pas à l'estimation de la variance de l'espérance conditionnelle mais à l'estimation de l'espérance de la variance conditionnelle. En effet, l'identité de l'analyse de la variance donne:

$$I_p = \Gamma + \Lambda$$

où $\Gamma := \text{var}(E[z|y])$ et $\Lambda := E[\text{cov}(z|y)]$. Les directions EDR standardisées, définies comme étant les K vecteurs propres associées aux K plus grandes valeurs propres de Γ , sont alors aussi les K vecteurs propres associés aux K plus petites valeurs propres de la matrice Λ .

En pratique, pour obtenir des estimations des directions EDR, il suffit de chercher des vecteurs propres d'une matrice qui estime Λ . Les estimateurs de Λ présentés par Hsing et Carroll (1992) sont basés sur les statistiques d'ordre $y_{(1)} \leq \dots \leq y_{(n)}$ et les valeurs $z_{(i)}$ concomitantes à $y_{(i)}$. Ces auteurs considèrent deux découpages possibles: (i) soit on fait des tranches "distinctes" de deux individus sur l'échantillon trié, (ii) soit on fait "glisser"

une tranche de deux individus sur ce même échantillon. La matrice Λ est alors estimée (i) soit par

$$\hat{\Lambda} = \frac{1}{n} \sum_{i=1}^{\lfloor n/2 \rfloor} (z_{(2i)*} - z_{(2i-1)*})(z_{(2i)*} - z_{(2i-1)*})' \quad (\text{cas des tranches "distinctes"})$$

où $\lfloor n/2 \rfloor$ désigne la partie entière de $n/2$, (ii) soit par

$$\tilde{\Lambda} = \frac{1}{2n} \sum_{i=2}^n (z_{(i)*} - z_{(i-1)*})(z_{(i)*} - z_{(i-1)*})' \quad (\text{cas des tranches "glissantes"}).$$

Remarque. Zhu et Ng (1995) ont proposé une généralisation de l'estimateur $\hat{\Lambda}$ (tranches distinctes de deux individus) en considérant le cas où les tranches contiennent non plus 2 individus mais un nombre c fixé arbitrairement dans chacune des H tranches. Leur estimateur a alors la forme suivante :

$$\begin{aligned} \hat{\Lambda} &= \frac{1}{H} \sum_{h=1}^H \left[\frac{1}{c-1} \sum_{j=1}^c (z_{(h,j)} - \frac{1}{c} \sum_{l=1}^c z_{(h,l)}) (z_{(h,j)} - \frac{1}{c} \sum_{l=1}^c z_{(h,l)})' \right] \\ &= \frac{1}{n(c-1)} \sum_{h=1}^H \left[\sum \sum_{1 \leq j \leq l \leq c} (z_{(h,l)} - z_{(h,j)}) (z_{(h,l)} - z_{(h,j)})' \right] \end{aligned}$$

où le double indiciage inférieur (h, j) se lit de la façon suivante : le premier indice h fait référence au numéro de la tranche et le second indice j fait référence au numéro d'ordre d'une observation dans cette tranche donnée. En pratique, l'utilisateur doit choisir le nombre c d'individus par tranche. Lorsque l'on travaille avec des petits échantillons, tout comme dans le cas de la méthode SIR classique, le choix de c semble avoir une certaine influence sur la qualité des estimations des directions EDR. Ainsi, cette généralisation n'apparaît pas très judicieuse à appliquer dans le cas d'échantillon de taille faible.

3.2.3 Régression inverse par lissage (Aragon et Saracco (1997))

Cette méthode alternative ne nécessite aucun découpage en tranches. Elle repose sur l'utilisation de la méthode des noyaux, méthode de régression fonctionnelle, pour estimer la matrice $\Gamma := \text{var}(E[z|y])$ lorsque y est une variable continue. Soit z_j la $j^{\text{ème}}$ composante de $z \in \mathbb{R}^p$. On considère les p fonctions ν_j de régression $y \mapsto \nu_j(y) := E[z_j|y]$. Etant donné que $E[z] = 0$, les composantes de Γ peuvent s'écrire en fonction des $\nu_j(y)$ et de la densité f de y . Plus précisément, on a pour tout $j = 1, \dots, p$ et pour tout $l = 1, \dots, p$,

$$\Gamma_{j,l} := \text{cov}(\nu_j(y), \nu_l(y)) = E[E[z_j|y]E[z_l|y]] = \int \nu_j(y)\nu_l(y)f(y)dy.$$

La matrice Γ est alors estimée composante par composante. Tout d'abord, pour estimer les p fonctions ν_j , $j = 1, \dots, p$, nous proposons d'utiliser la technique des estimateurs à noyaux de la régression de type Nadaraya-Watson sur les échantillons $\{(z_j)_i, y_i) \mid i = 1, \dots, n\}$. Pour quelques rappels sur la régression nonparamétrique par la méthode des noyaux ou pour plus de détails théoriques ou pratiques, voir par exemple Eubank (1988), Härdle (1990).

Plus précisément, soit $K_{h_j}(\cdot) = \frac{1}{h_j} K(\frac{\cdot}{h_j})$ un noyau où h_j est un réel positif appelé largeur de la fenêtre. L'estimateur $\hat{\nu}_j(y)$ de $\nu_j(y)$ est défini de la manière suivante :

$$\forall y \in \mathbb{R}, \hat{\nu}_j(y) = \frac{\sum_{i=1}^n (z_j)_i K_{h_{n,j}}(y - y_i)}{\sum_{i=1}^n K_{h_{n,j}}(y - y_i)}$$

où la largeur de la fenêtre optimale $h_{n,j}$ peut être obtenue de différentes manières, une des plus utilisées étant la méthode de validation croisée ("Cross Validation"). Pour estimer la densité f de y , nous utilisons l'estimateur à noyau de Parzen (1962) défini par :

$$\forall y \in \mathbb{R}, \hat{f}(y) = \frac{1}{n} \sum_{i=1}^n K_{h_n}(y - y_i)$$

où h_n est la largeur de fenêtre optimale (obtenue par exemple par validation croisée).

On obtient alors directement un estimateur des composantes de Γ de type "intégrale" :

$$\hat{\Gamma}_{j,l} = \int_{\mathbb{R}} \hat{\nu}_j(y) \hat{\nu}_l(y) \hat{f}(y) dy.$$

Un estimateur de type "somme" peut être aussi utilisé :

$$\tilde{\Gamma}_{j,l} = \frac{1}{n} \sum_{i=1}^n \hat{\nu}_j(y_i) \hat{\nu}_l(y_i) \mathbb{I}[\hat{f}(y_i) > b_n]$$

où b_n , appelé paramètre d'élagage ("trimming parameter"), est un réel positif dépendant de la taille n de l'échantillon. Le terme $\mathbb{I}[\hat{f}(y_i) > b_n]$ dans l'expression de $\tilde{\Gamma}_{j,l}$ a surtout un intérêt théorique. En pratique, le choix du paramètre d'élagage b_n ne pose pas de problème : on fixe une valeur pour b_n de manière à supprimer les individus de trop faible densité, c'est-à-dire les points aberrants. Très souvent, on conserve la totalité des observations.

Zhu et Fang (1996) ont aussi proposé une version de la régression inverse par lissage en utilisant un estimateur à noyaux de type "somme".

Remarque. Ces méthodes nécessitent un temps de calcul légèrement plus important par rapport aux autres méthodes présentées précédemment. En effet, quel que soit le choix de l'estimateur de Γ (somme ou intégrale), on est toujours amené à calculer les différents paramètres de lissage.

3.2.4 Quelques remarques sur les méthodes alternatives

Au vu des simulations réalisées (voir Aragon et Saracco (1997)), la méthode du Pooled Slicing (PSIR) et la méthode de régression inverse par lissage semblent se comporter mieux que les autres méthodes étudiées dans le cadre d'échantillon de taille n comprise entre 25 et 50 individus. Les estimateurs fondés sur les noyaux sont cependant légèrement plus coûteux en temps de calcul que les autres estimateurs. Les estimateurs de Hsing et Carroll sont simples à implémenter. Cependant sur des échantillons de petite taille, les estimations obtenues par des simulations sont souvent de mauvaise qualité. Pour terminer,

remarquons enfin que lorsque la variable à expliquer y est discrète, un unique découpage en tranches s'impose pour n'importe quelle taille d'échantillon : une tranche pour chaque valeur de y . On pourrait cependant dans certaines situations regrouper des tranches (ou modalités) de faible effectif.

4 Méthodes SIR multivariées (cas où $y \in \mathbb{R}^q$, $q > 1$)

Dans le cadre du modèle de référence multivarié ($M3$), Aragon, Li et Thomas-Agnan (1995) et Li, Aragon et Thomas-Agnan (1997) ont proposé différentes méthodes d'estimation de l'espace EDR. Pour un exposé introductif, on peut consulter Aragon (1997).

Ces diverses méthodes reposent sur des philosophies légèrement différentes, tout en utilisant toujours la propriété de caractérisation de l'espace EDR par la courbe de régression inverse et une étape de découpage au niveau de l'estimation. Dans cette section, nous présentons succinctement les idées, les aspects théoriques sous-jacents aux méthodes ainsi que leur mise en oeuvre sur un échantillon $\{(y_i, x_i), i = 1, \dots, n\}$.

La première méthode est fondée sur une utilisation directe de la propriété de la courbe de régression inverse. La seconde est une méthode qui agrège les diverses informations provenant des q composantes de y et enfin la dernière méthode repose sur l'observation d'une certaine dualité entre les directions EDR et les directions des variables les plus prédictives.

Deux applications des méthodes SIR multivariées sur des données réelles ont été faites. La première porte sur les distributions de revenus dans les communes de plus de 10000 habitants en France métropolitaine. Le traitement statistique de cette étude est détaillé dans Aragon, Li et Thomas-Agnan (1995). La seconde application concerne des données médicales (étude de la pression artérielle d'enfant), elle est décrite dans Li, Aragon et Thomas-Agnan (1997) et Aragon, Barthe, Cassadou et Thomas-Agnan (1996).

4.1 Complete Slicing et Marginal Slicing

De manière analogue au cas univarié (modèle ($M2*$)), si l'on considère une transformation T à valeurs dans \mathbb{R}^d avec $d \geq 1$, $T(y)$ satisfait un modèle ayant les mêmes indices $x'\eta_1, \dots, x'\eta_K$ que le modèle de régression semiparamétrique multivarié ($M3$) de référence. Ainsi le théorème de caractérisation de l'espace EDR standardisé et la conséquence sur la matrice Γ_T qui en découle, sont vérifiés et ceci quel que soit le choix de la transformation T . Différents choix intéressants sont alors possibles pour T en fonction du problème particulier que l'on est amené à traiter. Les méthodes du Complete Slicing et Marginal Slicing correspondent à des choix particulier de T .

Complete Slicing. Il s'agit de la méthode la plus directe. La transformation T correspond ici à un découpage de l'espace des y_i en un nombre H de tranches (hypercubes de \mathbb{R}^q) contenant à peu près le même nombre d'individus. Ainsi, T est à valeur dans $\{1, 2, \dots, H\}$. Il ne reste ensuite qu'à appliquer la méthode de régression inverse par tranches comme dans le cas univarié.

Le découpage en tranches est fait de manière récursive : on commence par découper selon la première composante de y puis selon sa seconde composante en tenant compte du découpage de la composante précédente, et ainsi de suite pour chacune des composantes. Le problème majeur qui apparaît ici est que le nombre de tranches augmente très rapidement avec la dimension de y . Ainsi, par exemple, pour $q = 5$, si l'on fait 3 tranches par composante, on obtiendra au total $3^5 = 243$ hypercubes, ce qui nécessite d'avoir au départ un échantillon de taille importante pour avoir un nombre raisonnable d'individus par tranche. Il apparaît alors avantageux de restreindre le nombre de composantes sur lesquelles on va faire un découpage en tranches, les composantes gardées étant celles qui résument ou conservent au mieux une bonne information contenue dans les y_i . Ce thème-là correspond à la méthode du Marginal Slicing présentée ci-après.

Remarque. Une autre manière de répartir en H groupes l'échantillon $\{y_i, i = 1, \dots, n\}$ est d'utiliser des méthodes de classification. Par exemple, les techniques non hiérarchiques de classification (comme la méthode des nuées dynamiques, voir Saporta (1990)) produisent directement une partition en un nombre fixé de classes (ou "tranches" dans le vocabulaire de SIR) mais on ne peut pas contrôler une bonne répartition des individus dans chacune des H classes.

Marginal Slicing. Cette méthode consiste à trouver une valeur scalaire $T(y)$ de y et à appliquer la méthode de régression inverse par tranches univariée sur l'échantillon $\{(T(y_i); x'_i) \mid i = 1, \dots, n\}$.

Le choix de la transformation T est très vaste et dépend en fait du contexte d'intérêt. Cela peut être une des composantes de y , ou bien la première composante ou les d premières composantes de l'analyse en composantes principales des y_i si le pourcentage de variance expliquée par ces premiers axes factoriels est suffisamment important. Dans le cas où les composantes de y mesurent le même phénomène (par exemple à des temps différents), T peut être ou bien la moyenne ou bien la médiane des q composantes de y . L'utilisation de la moyenne et de la médiane a été appliquée dans le cadre de l'étude sur la distribution des revenus de l'agglomération toulousaine (voir Aragon, Li et Thomas-Agnan (1995)). Il est intéressant de noter que l'on résume ici l'information de la variable à expliquer (composantes principales, médiane, moyenne, ...) sans tenir compte de la variable explicative x , ce qui n'est pas le cas des deux autres méthodes présentées aux sections 4.2 et 4.3.

Remarque. Quand la transformation $T(y)$ n'est plus scalaire, mais appartient à un sous-espace de dimension strictement inférieure à q (c'est le cas, par exemple, lorsque l'on retient plusieurs composantes principales), il suffit alors d'appliquer la méthode du Complete Slicing sur l'échantillon $\{(T(y_i)'; x'_i) \mid i = 1, \dots, n\}$ à la place de la méthode SIR univariée.

4.2 Pooled Marginal Slicing

Cette méthode repose sur l'observation suivante. Etant données deux transformations à valeurs réelles T_1 et T_2 de y ainsi que les matrices de covariances associées

$$\Gamma_1 = \text{var}(E[z|T_1(y)]) \text{ et } \Gamma_2 = \text{var}(E[z|T_2(y)]),$$

on peut construire une combinaison convexe Γ^P de Γ_1 et Γ_2

$$\Gamma^P = w_1\Gamma_1 + w_2\Gamma_2$$

avec w_1 et w_2 deux poids positifs tels que $w_1 + w_2 = 1$. De manière analogue au cas univarié, les K vecteurs propres associés aux K valeurs propres non nulles de Γ^P sont des directions EDR standardisées. Ceci se généralise sans aucun problème au cas où l'on considère plus de deux transformations T .

La méthode du Pooled Marginal Slicing utilise q transformations $T_j(y) = t_j(y_j)$ qui sont des discrétisations en H_j tranches de chacune des q composantes y_j de y . On définit alors la matrice agrégée :

$$\Gamma^P = \sum_{j=1}^q w_j \Gamma_j$$

où $\Gamma_j = \text{var}(E[z|t_j(y_j)])$ et les w_j sont une suite de poids positifs tels que $\sum_{j=1}^q w_j = 1$.

Pour la version empirique du Pooled Marginal Slicing, chaque matrice Γ_j est estimée par $\hat{\Gamma}_j$ de la même manière que dans le cas univarié. Pour le choix des poids w_j , on peut soit les prendre tous égaux, soit les prendre égaux à la somme des K plus grandes valeurs propres de $\hat{\Gamma}_j^s$ que l'on normalisera ensuite à 1. Une estimation de Γ^P est alors :

$$\hat{\Gamma}^P = \sum_{j=1}^q w_j \hat{\Gamma}_j$$

Les directions EDR standardisées estimées sont alors les K vecteurs propres associés aux K plus grandes valeurs propres de la matrice $\hat{\Gamma}^P$.

4.3 Alternating SIR

Li, Aragon et Thomas-Agnan (1997) proposent une théorie concernant les variables les plus prévisibles (Most Predictable) en s'inspirant des travaux de Hotteling (1935, 1936). Leur théorie (contrairement à celle d'Hotteling) utilise des prédicteurs non linéaires et se ramène à une décomposition aux valeurs propres du même type que celle de SIR : il suffit en effet simplement d'invertir les rôles de x et de y . La méthode Alternating SIR repose sur l'observation de cette dualité entre les directions EDR (combinaisons linéaires de x) et les directions des variables les plus prévisibles (combinaisons linéaires de y).

Au niveau de la mise en oeuvre pratique, l'utilisation de cette dualité revient à utiliser un algorithme qui à chaque itération calcule les directions EDR et les directions les plus prévisibles. On arrête les itérations lorsque la ou les directions retenues se stabilisent. La convergence de l'algorithme en un nombre fini d'étapes a été montré dans Li, Aragon et Thomas-Agnan (1995). On en trouve une mise en oeuvre en Gauss dans Aragon (1997).

Références bibliographiques

- ARAGON, Y. (1997). A Gauss implementation of Multivariate Sliced Inverse Regression. *Computational Statistics* 12 355-372.
- ARAGON, Y., BARTHE, PH., CASSADOU, S. and THOMAS-AGNAN, CH. (1996). Analysing ambulatory blood pressure monitoring data with Multivariate Sliced Inverse Regression. *Rapport technique du GREMAQ, Université des Sciences Sociales de Toulouse.*
- ARAGON, Y., LI, K. C. and THOMAS-AGNAN, CH. (1995). Modelling income distributions using Multivariate Sliced Inverse Regression. *Rapport technique du GREMAQ, Université des Sciences Sociales de Toulouse.*
- ARAGON, Y. and SARACCO, J. (1997). Sliced Inverse Regression (SIR): an appraisal of small sample alternatives to slicing. *Computational Statistics* 12 109-130.
- BRILLINGER, D. R. (1983). A generalized linear model with "gaussian" regressor variables. *Festschrift for Erich L. Lehmann in Honor of His Sixty-Fifth Birthday (P. J. Bickel, K. A. Doksum, and J. L. Hodges, eds.)* 97-114 Wadsworth, Belmont, Calif.
- CARROLL, R. J. and LI, K. C. (1992). Measurement error regression with unknown link: dimension reduction and data visualization. *Journal of the American Statistical Association* 87 1040-1050.
- COOK, R. D. and NACHTSHEIM, C. J. (1994). Re-weighting to achieve elliptically contoured covariates in regression. *Journal of the American Statistical Association* 89 592-599.
- COOK, R. D. and WIESBERG, S. (1991). Discussion of "Sliced Inverse Regression". *Journal of the American Statistical Association* 86 328-332.
- DUAN, N. and LI, K. C. (1991). Slicing regression: a link-free regression method. *The Annals of Statistics* 19 505-530.
- EUBANK, R. (1988). *Spline smoothing and nonparametric regression*. New York: Dekker.
- FERRÉ, L. (1996). Choix de dimension en régression inverse par tranches (SIR). *C. R. Acad. Sci. Paris, t. 323, série I, n° 4*, 403-406.
- FERRÉ, L. (1998). Determining the dimension in SIR and related methods. *Journal of the American Statistical Association* 441 132-140.
- GOURIÉROUX, C. (1989). *Econométrie des variables qualitatives*. Economica.
- HAERDLE, W. (1990). *Applied nonparametric regression*. Cambridge University Press.
- HALL, P. (1989). On projection pursuit regression. *The Annals of Statistics* 17 589-605.
- HALL, P. and LI, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics* 21 867-889.
- HSING, T. and CARROLL, R. J. (1992). An asymptotic theory for Sliced Inverse regression. *The Annals of Statistics* 20 1040-1061.
- LI, K. C. (1991). Sliced inverse regression for dimension reduction, with discussions. *Journal of the American Statistical Association* 86 316-342.
- LI, K. C., ARAGON, Y. and THOMAS-AGNAN, CH. (1997). Analysis of Multivariate outcome data: SIR and a nonlinear theory of Hotelling's most predictable variates. *Soumis à Journal of the American Statistical Association.*
- PARZEN, E. (1962). On estimation of probability density function and mode. *Ann. Math. Statist.* 16 629-647.
- SAPORTA, G. (1990). *Probabilités, Analyse des Données et Statistique*. Editions Technip.

- SARACCO, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Communications in statistics - Theory and methods* **26** 2141-2171.
- SARACCO, J. (1998). Implémentation en Splus des méthodes SIR univariées et multivariées. *La revue de Modulad (même numéro)*.
- SCHOTT, J. R. (1994). Determining the dimensionality in Sliced Inverse Regression. *Journal of the American Statistical Association* **89** 141-148.
- ZHU L. X. and NG, K. W. (1995). Asymptotics of sliced inverse regression. *Statistica Sinica* **5** 727-736.
- ZHU L. X. and FANG, K. T. (1996). Asymptotics for kernel estimate of Sliced Inverse Regression. *The Annals of Statistics* **24** 1053-1068.

