

Une méthode de régression inverse par tranches pour expliquer une variable multivariée

Raphaël Coudret^(1,2), Stéphane Girard⁽³⁾, Jérôme Saracco^(1,2)

⁽¹⁾Equipe CQFD, INRIA Bordeaux - Sud-Ouest

⁽²⁾ EPOC, IMB et IPB, Université de Bordeaux

⁽³⁾ Equipe Mistis, INRIA Grenoble - Rhône-Alpes et Laboratoire Jean Kuntzmann

5^{èmes} Rencontres des Jeunes Statisticiens,
mardi 27 août 2013, Aussois

Sommaire

La régression inverse par tranches (SIR)

Adaptations pour une variable multivariée

Conclusion et perspectives

La régression inverse par tranches (SIR)

Un modèle semi-paramétrique

Définitions et notations

- ▶ Nous appelons $\text{Vect}(\mathbf{B})$ **espace EDR**, pour *efficient dimension reduction*.
- ▶ Notons β_1, \dots, β_K les colonnes de \mathbf{B} . Ce sont des **directions EDR**.
- ▶ Notons $\{\mathbf{x}_i\}_{i \in \{1, \dots, n\}}$ un échantillon i.i.d. de même loi que \mathbf{x} .
- ▶ Pour chaque $j \in \{1, \dots, q\}$, $\{y_i^{(j)}\}_{i \in \{1, \dots, n\}}$ est un échantillon i.i.d. de même loi que $y^{(j)}$.

Procédure d'estimation lorsque $q = 1$

Considérons l'exemple suivant :

$$y^{(1)} = (\mathbf{x}'\beta_1)^3 + \varepsilon^{(1)}$$

avec

- ▶ $\mathbf{x} = (x^{(1)}, x^{(2)})' \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,
- ▶ $\boldsymbol{\mu} = (0, 1)'$, $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$,
- ▶ $\varepsilon^{(1)} \sim \mathcal{N}(-2, 2500)$,
- ▶ $\beta_1 = (1, -3)'$.

Procédure d'estimation lorsque $q = 1$

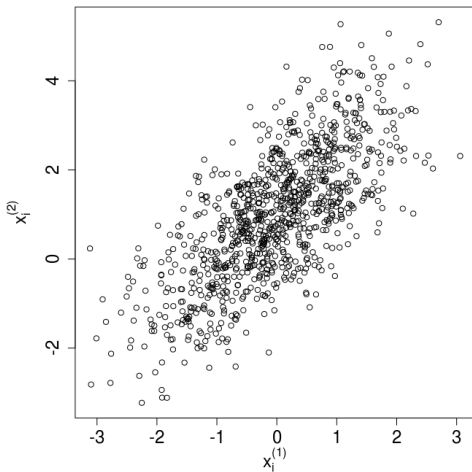


Figure : Simulation d'un échantillon $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ avec $n = 1000$.

Procédure d'estimation lorsque $q = 1$

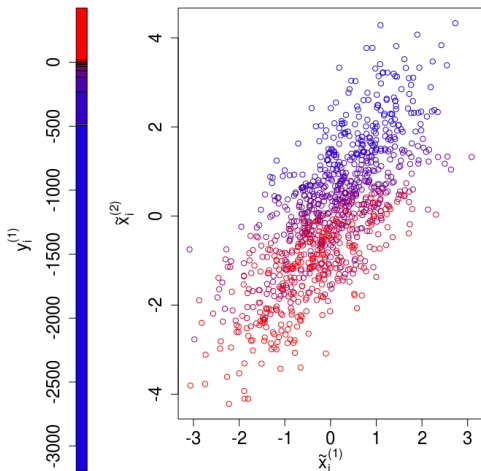


Figure : Nous calculons, pour tout $i \in \{1, \dots, n\}$, $\tilde{x}_i = x_i - \hat{\mu}$ où $\hat{\mu}$ est l'estimateur usuel de μ . Nous découpons \mathbb{R} en $H = 10$ intervalles disjoints (ou tranches) tels qu'il y ait n/H éléments de $\{y_i^{(1)}\}_{i \in \{1, \dots, n\}}$ dans chacune d'elles. Nous appliquons ce tranchage à $\{\tilde{x}_i\}_{i \in \{1, \dots, n\}}$.

Procédure d'estimation lorsque $q = 1$

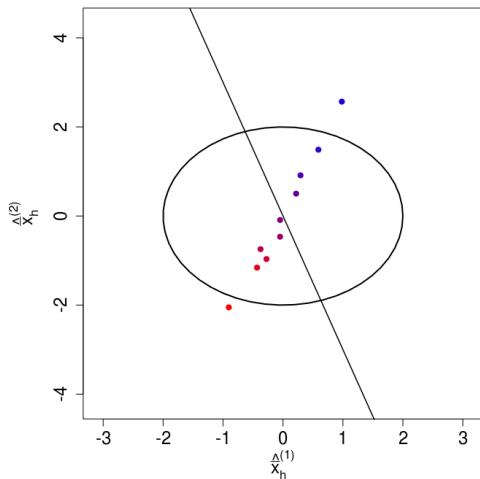


Figure : Pour tout $h \in \{1, \dots, H\}$, nous calculons le barycentre $\hat{\tilde{x}}_h$ des points $\{\tilde{x}_i\}_{i \in \{1, \dots, n\}}$ appartenant à la tranche h : on ne retrouve pas la direction EDR (ligne noire).

Procédure d'estimation lorsque $q = 1$

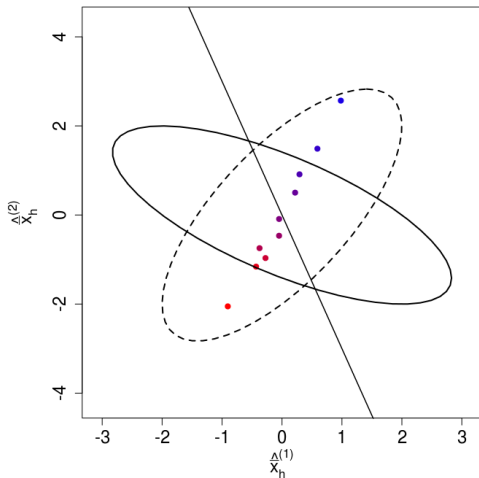


Figure : Les points ont tendance à se répartir dans une ellipse comme celle en pointillés. Pour retirer cette tendance, nous cherchons $\hat{\beta}_1$ parmi les vecteurs tels que $\hat{\beta}_1' \Sigma \hat{\beta}_1$ soit constant (ellipse en trait plein).

Procédure d'estimation lorsque $q = 1$

Soit $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_H)$, $\hat{\mathbf{M}} = \hat{\mathbf{x}}\hat{\mathbf{x}}'$ et

$$\hat{\beta}_1 = \arg \max_u \frac{u' \hat{\mathbf{M}} u}{u' \hat{\Sigma} u}$$

$\hat{\beta}_1$ est le “premier” **vecteur propre généralisé** de $\hat{\mathbf{M}}$ et $\hat{\Sigma}$.

En pratique, nous n'avons pas Σ , nous utilisons l'estimateur usuel $\hat{\Sigma}$ à la place.

Lorsque $\hat{\Sigma}$ est inversible, $\hat{\beta}_1$ est également le “premier” vecteur propre de $\hat{\Sigma}^{-1} \hat{\mathbf{M}}$.

Dans notre exemple, nous trouvons $\hat{\beta}_1 = (0.30, -0.95)'$.

Procédure d'estimation lorsque $q = 1$

Pour tout $h \in \{1, \dots, H\}$, soient s_h la tranche h et,

$$\bar{\mathbf{x}}_h = \mathbb{E}(\mathbf{x}|y^{(1)} \in s_h), \text{ et soient } \bar{\mathbf{x}} = (\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_H), \text{ et } \mathbf{M} = \bar{\mathbf{x}}\bar{\mathbf{x}}'.$$

Soit $\hat{\mathbf{B}}$ la matrice contenant les K "premiers" vecteurs propres de $\hat{\Sigma}^{-1}\hat{\mathbf{M}}$.

Si nous avons la **condition de linéarité** suivante pour tout $b \in \mathbb{R}^p$,

$$\exists(c_0, \dots, c_K) \in \mathbb{R}^{K+1}, \mathbb{E}(\mathbf{x}'b|\mathbf{x}'\beta_1, \dots, \mathbf{x}'\beta_K) = c_0 + \sum_{k=1}^K c_k \mathbf{x}'\beta_k$$

et si \mathbf{M} a K valeurs propres strictement positives alors $\text{Vect}(\hat{\mathbf{B}})$ converge en probabilité vers $\text{Vect}(\mathbf{B})$.

Approche introduite par Li (1991). Voir aussi Zhu et Ng (1995).

Adaptations pour une variable multivariée

Estimation avec MSIR

Complexifions le modèle précédent :

$$\begin{cases} y^{(1)} &= (\mathbf{x}'\beta_1)^3 + \varepsilon^{(1)}, \\ y^{(2)} &= 10 (\mathbf{x}'\beta_1)^2 + 100 \mathbb{I}(\mathbf{x}'\beta_1 > 0) + \varepsilon^{(2)}, \end{cases}$$

avec $\varepsilon^{(2)} \sim \mathcal{N}(-2, 2500)$.

Après génération des éléments de $\{y_i^{(2)}\}_{i \in \{1, \dots, n\}}$, nous réalisons un second tranchage de $\{\tilde{\mathbf{x}}_i\}_{i \in \{1, \dots, n\}}$.

Pour chaque tranche, nous obtenons un barycentre correspondant à la variable $y^{(2)}$.

Estimation avec MSIR

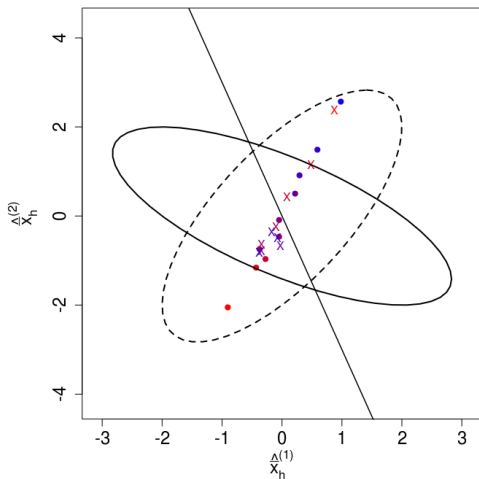


Figure : Barycentres relatifs à $y^{(1)}$ (points) et à $y^{(2)}$ (croix).

Estimation avec MSIR

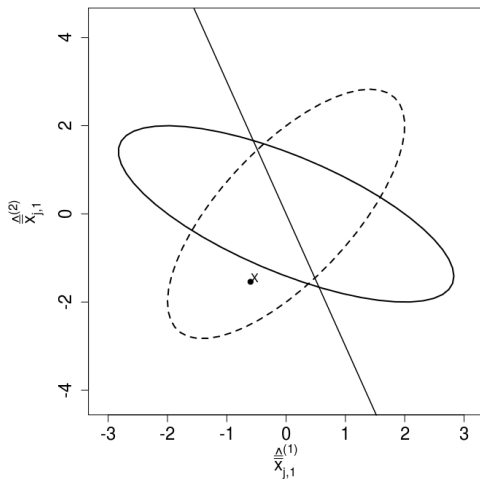


Figure : Pour chaque $y \in \{1, \dots, q\}$, nous remplaçons les barycentres correspondants par K points $\hat{\hat{x}}_{j,1}, \dots, \hat{\hat{x}}_{j,K}$ tels que l'estimation de l'espace EDR reste inchangée.

Estimation avec MSIR

Soit $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_{1,1}, \dots, \hat{\mathbf{x}}_{1,K}, \dots, \hat{\mathbf{x}}_{q,1}, \dots, \hat{\mathbf{x}}_{q,K})$ et $\hat{\mathbf{M}} = \hat{\mathbf{x}}\hat{\mathbf{x}}'$.

Nous estimons l'espace EDR par celui engendré par la matrice $\hat{\mathbf{V}}$ des K "premiers" vecteurs propres de la matrice $\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{M}}$.

Si l'estimation univariée de l'espace EDR à partir de $y^{(j)}$ mène à une matrice notée $\hat{\mathbf{B}}^{(j)}$, nous avons $(\hat{\mathbf{x}}_{j,1}, \dots, \hat{\mathbf{x}}_{j,K}) = \hat{\mathbf{\Sigma}}\hat{\mathbf{B}}^{(j)}$.

Pour chaque $j \in \{1, \dots, q\}$, nous définissons $\mathbf{M}^{(j)}$ à partir de $y^{(j)}$ comme nous avons défini \mathbf{M} à partir de $y^{(1)}$.

Sous la condition de linéarité et si pour tout $j \in \{1, \dots, q\}$, $\mathbf{M}^{(j)}$ a K valeurs propres strictement positives alors $\text{Vect}(\hat{\mathbf{V}})$ converge en probabilité vers $\text{Vect}(\mathbf{B})$ (Coudret, Girard et Saracco).

Estimation avec MSIR

Si nous réalisons l'estimation directement à partir des qH barycentres, nous obtenons le *pooled marginal slicing* (PMS). Voir Aragon (1997), Saracco (2005).

Une autre méthode de fabrication des barycentres mène au *k-means inverse regression* (KIR) de Setodji et Cook (2004).

MSIR, PMS et KIR offrent des résultats similaires en simulation avec un léger avantage pour PMS.

Détails

Pourquoi utiliser MSIR plutôt que PMS ?

Classification

Nous pouvons être amenés à essayer d'expliquer des variables qui ne partagent pas le même espace EDR de dimension K .

Nous voulons créer des groupes de variables sur lesquels appliquer notre modèle.

Pour deux variables $y^{(j_1)}$ et $y^{(j_2)}$, soient $\text{Vect}(\mathbf{B}^{(j_1)})$ et $\text{Vect}(\mathbf{B}^{(j_2)})$ les espaces EDR de dimension K qui leur sont respectivement associés.

Si $\text{Vect}(\mathbf{B}^{(j_1)}) = \text{Vect}(\mathbf{B}^{(j_2)})$ alors $\forall \varepsilon > 0$,

$$\mathbb{P} \left(\left| \frac{1}{K} \left\| \widehat{\mathbf{B}}^{(j_1)} \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{B}}^{(j_2)} \right\|_2^2 - 1 \right| > \varepsilon \right) \rightarrow 0$$

lorsque $n \rightarrow \infty$.

Classification

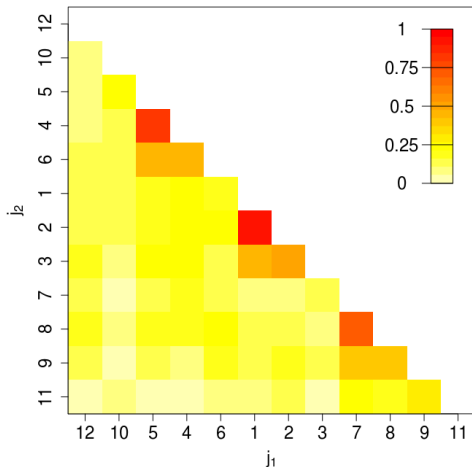


Figure : Valeurs de $\frac{1}{K} \left\| \widehat{\mathbf{B}}(j_1) \widehat{\Sigma} \widehat{\mathbf{B}}(j_2) \right\|_2^2$ pour un modèle à 12 variables réparties en groupes.

[Voir le modèle](#)

Classification

En utilisant uniquement $\{\widehat{\mathbf{B}}^{(j)}\}_{j \in \{1, \dots, q\}}$ et $\widehat{\Sigma}$ nous pouvons

- ▶ calculer le critère pour classifier les variables,
- ▶ estimer les espaces EDR de chacun des groupes.

Faire de même avec la méthode PMS nécessite au moins le stockage des barycentres soit H points par variable à la place de K avec MSIR.

En pratique K est souvent proche de 1 alors que H doit être assez grand pour obtenir des estimations précises des espaces EDR.

Conclusion et perspectives

Conclusion et perspectives

- ▶ Utilisation de MSIR à la place de PMS :
 - performances légèrement diminuées (estimation),
 - coûts moindres en ressources de calcul (classification).
- ▶ En pondérant $\widehat{\Sigma B}^{(1)}, \dots, \widehat{\Sigma B}^{(q)}$ par les valeurs propres associées à chaque SIR univarié, nous obtenons de meilleurs résultats.
- ▶ Applications à des données réelles :
 - notes des élèves des écoles de Minneapolis,
 - données spectrales de la surface de Mars ($q = 352$).
- ▶ Applications possibles à des données génétiques :
 - plusieurs marqueurs, plusieurs phénotypes,
 - développement de tests d'hypothèse.

Références

Y. Aragon.

A Gauss implementation of multivariate sliced inverse regression.
Computational Statistics, 12 :355–372, 1997.

R. Coudret, S. Girard, and J. Saracco.

A new sliced inverse regression method for multivariate response regression.
En revision.

L. Ferré.

Determining the dimension in sliced inverse regression and related methods.
Journal of the American Statistical Association, 93 :132–140, 1998.

K.-C. Li.

Sliced inverse regression for dimension reduction.
Journal of the American Statistical Association, 86(414) :316–327, 1991.

J. Saracco.

Asymptotics for pooled marginal slicing estimator based on SIR_α approach.
Journal of Multivariate Analysis, 96 :117–135, 2005.

C. M. Setodji and R. D. Cook.

K-means inverse regression.
Technometrics, 46 :421–429, 2004.

L.-X. Zhu and K.-W. Ng.

Asymptotics of sliced inverse regression.
Statistica Sinica, 5 :727–736, 1995.

Simulations

Nous générons des échantillons de taille $n = 100$ à partir du modèle suivant :

$$\begin{cases} y^{(1)} &= \mathbf{x}'\beta_1 + \varepsilon^{(1)}, \\ y^{(2)} &= (\mathbf{x}'\beta_1)^3 + 3\varepsilon^{(2)}, \\ y^{(3)} &= \mathbf{x}'\beta_1(1 + \varepsilon^{(3)}), \end{cases} \quad (4.1)$$

où :

- ▶ $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,
- ▶ $\boldsymbol{\mu} \sim \mathcal{N}((0, \dots, 0), \mathbf{I}_{20})$,
- ▶ $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + 0.1\mathbf{I}_{20}$,
- ▶ \mathbf{L} est une matrice 20×20 et $\text{vec}(\mathbf{L}) \sim \mathcal{N}((0, \dots, 0), \mathbf{I}_{400})$,
- ▶ $\beta_1 = (1, 2, 3, 4, 5, 1, \dots, 1)$,
- ▶ pour tout $j \in \{1, \dots, 3\}$, $\varepsilon^{(j)} \sim \mathcal{N}(0, 1)$,

Simulations

Nous estimons $\text{Vect}(\beta_1)$ avec

- ▶ SIR univarié pour chaque variable ($\widehat{\mathbf{B}}^{(1)}$, $\widehat{\mathbf{B}}^{(2)}$, $\widehat{\mathbf{B}}^{(3)}$),
- ▶ la version multivariée de SIR présentée ici ($\widehat{\mathbf{V}}$),
- ▶ une version alternative utilisant une pondération ($\widehat{\widetilde{\mathbf{V}}}$),
- ▶ le *k-means inverse regression* ($\widehat{\mathbf{V}}_{\text{KIR}}$),
- ▶ le *pooled marginal slicing* ($\widehat{\mathbf{V}}_{\text{PMS}}$).

Le critère de qualité de l'estimation choisi est le *squared trace correlation*. Voir par exemple Ferré (1998).

Simulations

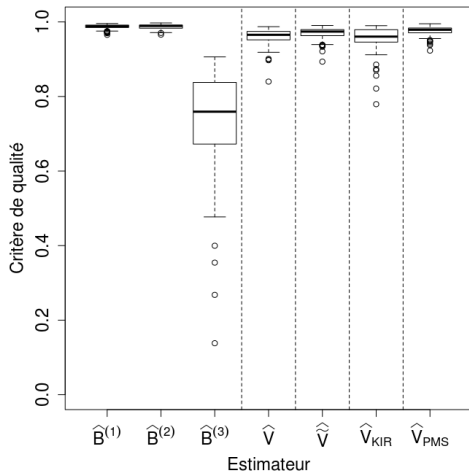


Figure : Boîtes à moustaches de 100 critères de qualité pour chaque estimateur appliqué sur des données simulées à partir du modèle (4.1).

Modèle avec plusieurs espaces EDR

$$\left\{ \begin{array}{l} y^{(1)} = \mathbf{x}'\beta_1 + \varepsilon^{(1)}, \\ y^{(2)} = (\mathbf{x}'\beta_1)^3 + 3\varepsilon^{(2)}, \\ y^{(3)} = \mathbf{x}'\beta_1(1 + \varepsilon^{(3)}), \\ \\ y^{(4)} = \mathbf{x}'\gamma_1 + \varepsilon^{(4)}, \\ y^{(5)} = (\mathbf{x}'\gamma_1)^3 + 3\varepsilon^{(5)}, \\ y^{(6)} = \mathbf{x}'\gamma_1(1 + \varepsilon^{(6)}), \\ \\ y^{(7)} = \mathbf{x}'\delta_1 + \varepsilon^{(7)}, \\ y^{(8)} = (\mathbf{x}'\delta_1)^3 + 3\varepsilon^{(8)}, \\ y^{(9)} = \mathbf{x}'\delta_1(1 + \varepsilon^{(9)}), \\ \\ y^{(10)} = \varepsilon^{(10)}, \\ y^{(11)} = \varepsilon^{(11)}, \\ y^{(12)} = \varepsilon^{(12)}, \end{array} \right.$$

Modèle avec plusieurs espaces EDR

où :

- ▶ $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,
- ▶ $\boldsymbol{\mu} \sim \mathcal{N}((0, \dots, 0), \mathbf{I}_{20})$,
- ▶ $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + 0.1\mathbf{I}_{20}$,
- ▶ \mathbf{L} est une matrice 20×20 et $\text{vec}(\mathbf{L}) \sim \mathcal{N}((0, \dots, 0), \mathbf{I}_{400})$,
- ▶ $\beta_1 = (1, 2, 3, 4, 5, 1, \dots, 1)$,
- ▶ $\gamma_1 = (5, 4, 3, 2, 1, \dots, 5, 4, 3, 2, 1)$,
- ▶ $\delta_1 = (1, -1, 2, -2, 1, -1, \dots, 1, -1)$,
- ▶ pour tout $j \in \{1, \dots, 12\}$, $\varepsilon^{(j)} \sim \mathcal{N}(0, 1)$,
- ▶ $n = 1000$.