
PUBLICATIONS DU GROUPE DE TRAVAIL STAPH:
STATISTIQUE FONCTIONNELLE ET OPÉRATORIELLE

STAPH-2009-01 *Recueil de résumés de l'année 2008-09*

ALAIN BOUDOU, FRÉDÉRIC FERRATY, YVES ROMAIN, PASCAL SARDA, PHILIPPE VIEU
ET SYLVIE VIGUIER-PLA

Institut de Mathématiques, Université Paul Sabatier, Toulouse, France

Presmoothing and functional linear regression

MARTÍNEZ-CALVO* Adela

* Adresse pour correspondance:
Department of Statistics and OR
Faculty of Mathematics
University of Santiago de Compostela
CP 15782, Santiago de Compostela, Spain
e-mail: adela.martinez@usc.es

Résumé

Nowadays the progress of computing and measure tools allows us to have access to data that can be observed in such a fine grid that they can be seen as discretized observations of a functional variable (see Ramsay and Silverman [13], Ramsay and Silverman [14] or Ferraty and Vieu [9], for examples). In these cases, multivariate regression techniques are not adequate due to large number of explanatory variables and high correlation between them. For this reason the classical regression models have been adapted to the functional context where the response Y and/or the explanatory variable X are valued in a functional space (see Ramsay and Silverman [14], for a *parametric* state of art and Ferraty and Vieu [9], for a *nonparametric* one). Among the functional regression models, the functional linear model with scalar response have been the subject of numerous studies.

In order to make everything formal, let $(E, \langle \cdot, \cdot \rangle)$ be a real separable Hilbert space and let $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$ be its associated norm (that includes the multivariate case, $E = \mathbb{R}^p$). Let us consider the functional linear model given by

$$Y = m(X) + \epsilon = \langle X, \theta \rangle + \epsilon,$$

where Y is a real random variable, X is a random variable valued in E such that $\mathbb{E}(X) = 0$ and $\mathbb{E}(\|X\|^2) < \infty$, θ is a parameter valued in E that satisfies $\|\theta\|^2 < \infty$, and ϵ is a real random variable that verifies $\mathbb{E}(\epsilon) = 0$, $\mathbb{E}(\epsilon^2) = \sigma^2$ and $\mathbb{E}(\epsilon X) = 0$. Many authors have proposed procedures for estimating the model parameter θ , being one of the most habitual methods the basis function systems as Fourier series, wavelets or splines (see a general

revision in Ramsay and Silverman [14], and see Cardot *et al.* [4] or Crambes *et al.* [6] for the splines approach). Another well-known methodology is based on functional principal components analysis (FPCA), and has been developed and analysed in Cardot *et al.* [3], Cardot *et al.* [4], Cai and Hall [2], Hall and Hosseini-Nasab [11], Hall and Horowitz [10] or Cardot *et al.* [5]. We have revisited the FPCA estimator in order to improve its behavior in terms of the conditional mean-squared error by introducing presmoothing techniques.

The reason why we have decided to introduce presmoothing methods comes from the eighties when Faraldo-Roca and González-Manteiga [8] and Cristóbal-Cristóbal *et al.* [7] proposed, for real case and multivariate case respectively, the application of the least squares principle on the pairs $(X_i, \hat{m}(X_i))$, where \hat{m} is a nonparametric kernel-type estimation of the regression function m , instead of (X_i, Y_i) to reduce the mean-square error of the classical least-squares estimates for the linear regression model. The efficiency of these estimates was stated, except compact support case (later, Janssen *et al.* [12] have shown that the inefficiency problem in the compact support case can be rectified using *boundary kernels*). Furthermore, a similar procedure was developed by Akritas [1] to fit polynomial regression models to data with incomplete observations. We have realized that estimates exposed in Cristóbal-Cristóbal *et al.* [7] can be seen as *presmoothed versions* of the estimates based on multivariate PCA. Hence we are interested in introducing presmoothing on the FPCA estimator to check if we can reduce the conditional mean-squared error.

References

- [1] M. G. Akritas, *On the use of nonparametric regression techniques for fitting parametric regression models*, Biometrics **52** (1996), no. 4, 1342–1362.
- [2] T.T. Cai and P. Hall, *Prediction in functional linear regression*, Annals of Statistics **34** (2006), no. 5, 2159–2179.
- [3] H. Cardot, F. Ferraty, and P. Sarda, *Functional linear model*, Statistics and Probability Letters **45** (1999), no. 1, 11–22.
- [4] H. Cardot, F. Ferraty, and P. Sarda, *Spline estimators for the functional linear model*, Statistica Sinica **13** (2003), no. 3, 571–591.
- [5] H. Cardot, A. Mas, and P. Sarda, *Clt in functional linear regression models*, Probability Theory and Related Fields **138** (2007), no. 3-4, 325–361.

- [6] C. Crambes, A. Kneip, and P. Sarda, *Smoothing splines estimators for functional linear regression*, *Annals of Statistics* (2008), To appear.
- [7] J. A. Cristóbal-Cristóbal, P. Faraldo-Roca, and W. González-Manteiga, *A class of linear regression parameter estimators constructed by nonparametric estimation*, *Annals of Statistics* **15** (1987), no. 2, 603–609.
- [8] P. Faraldo-Roca and W. González-Manteiga, *On efficiency of a new class of linear regression estimates obtained by preliminary non-parametric estimation*, *New Perspectives in Theoretical and Applied Statistics* (New York) (M. Puri, ed.), Wiley, 1985, pp. 229–242.
- [9] F. Ferraty and P. Vieu, *Nonparametric functional data analysis : theory and practice*, Springer, New York, 2006.
- [10] P. Hall and J.L. Horowitz, *Methodology and convergence rates for functional linear regression*, *Annals of Statistics* **35** (2007), no. 1, 70–91.
- [11] P. Hall and M. Hosseini-Nasab, *On properties of functional principal components analysis*, *Journal of the Royal Statistical Society: Series B* **68** (2006), no. 1, 109–126.
- [12] P. Janssen, Swanepoel J., and N. Veraverbeke, *Efficiency of linear regression estimators based on presmoothing*, *Communications in Statistics - Theory and Methods* **30** (2001), no. 10, 2079–2097.
- [13] J.O. Ramsay and B.W. Silverman, *Applied functional data analysis. Methods and case studies*, Springer, New York, 2002.
- [14] J.O. Ramsay and B.W. Silverman, *Functional data analysis*, Second ed., Springer, New York, 2005.

Normality test based on nonlinear principal components

GOIA Aldo*, **SALINELLI, Ernesto** and **SARDA, Pascal**

* Adresse pour correspondance:

Università del Piemonte Orientale, Dipartimento SEMeQ,

Via Perrone, 18 - 28100 NOVARA (ITALY)

e-mail: aldo.goia@eco.unipmn.it

Résumé

Consider an absolutely continuous r.v. X with zero mean, finite variance σ^2 and density f_X , and $\dot{W}_{f_X}^{1,2}$ the weighted Sobolev space of centered, square integrable functions with first derivative square integrable too. In Goia and Salinelli (2007) a definition of Nonlinear Principal Components (NLPCs) for X has been introduced specializing to the univariate case the variance maximizing problem used to define NLPCs for a random vector in Salinelli (1998). More precisely, we say that the r.v. $Z_j = \varphi_j(X)$ is the j -th nonlinear principal component of X if φ_j is a solution of the problem

$$\begin{aligned} \max \quad & \mathbb{E} [u(X)^2] \\ \text{sub} \quad & \mathbb{E} [u'(X)^2] = 1 \quad u \in \dot{W}_{f_X}^{1,2} \\ & \mathbb{E} [u(X) \varphi_s(X)] = 0 \quad s = 1, 2, \dots, j-1, \quad j > 1. \end{aligned} \quad (0.1)$$

When $X \sim \mathcal{N}(0, \sigma^2)$ the problem (0.1) admits the solutions (see Goia and Salinelli, 2007)

$$\varphi_j(x) = (j\sigma^{2(j-1)})^{-1/2} H_j(x) \quad j = 1, 2, \dots$$

where H_j is the j -th Hermite polynomial defined by

$$H_j(x) = \frac{(-\sigma^2)^j}{\sqrt{j!}} e^{x^2/2\sigma^2} \frac{d^j}{dx^j} \left(e^{-x^2/2\sigma^2} \right) \quad j = 1, 2, \dots$$

The variances λ_j of the NLPCs $Z_j = \varphi_j(X)$ are expressed in a simple form in terms of the variance of X :

$$\lambda_j = \frac{\sigma^2}{j} \quad j = 1, 2, \dots$$

The first NLPC transformation is the identity function $\varphi_1(x) = x$ and, as proved in Salinelli (2008) in a multivariate framework, this happens only in the gaussian case. This result implies that univariate normality is characterized only by the equality $\lambda_1 = \sigma^2$.

We use this fine property to derive a test of normality: consider the following statistical hypothesis

$$\mathcal{H}_0 : X \sim \mathcal{N}(0, \sigma^2) \quad \text{vs.} \quad \mathcal{H}_1 : X \sim \mathcal{L}(0, \sigma^2)$$

where σ^2 is known, and \mathcal{L} is a not gaussian distribution that belongs to the class of probability laws for which NLPCs exist. By the characterization property of the first NLPC this hypothesis is equivalent to

$$\mathcal{H}_0 : \lambda_1 = \sigma^2 \quad \text{vs.} \quad \mathcal{H}_1 : \lambda_1 > \sigma^2$$

This allows to define the test statistic

$$\sqrt{\frac{n}{2}} \frac{\widehat{\lambda}_{1,n} - \sigma^2}{\sigma^2}$$

where $\widehat{\lambda}_{1,n}$ is an estimator of λ_1 which is based on orthogonal polynomials, which is computed on a sample $\{X_i, i = 1, \dots, n\}$ of i.i.d. r.v.s drawn from X . The asymptotic distribution of this statistic is derived.

Since in practice it is more interesting to consider the composite null hypothesis with both mean and variance unknown, we adapt the test statistic in this way

$$\delta_n^* = \sqrt{n} \frac{\widehat{\lambda}_{1,n} - \widehat{\sigma}^2}{\widehat{\sigma}^2}$$

where $\widehat{\lambda}_{1,n}$ is computed on the centered sample $\{(X_i - \overline{X}), i = 1, \dots, n\}$, being \overline{X} the sample mean and $\widehat{\sigma}^2$ the empirical estimator of the variance.

The finite sample performances of the test based on δ_n^* are investigated by an extensive Monte Carlo study: under some different experimental conditions we analyze the level and the power and we compare the results with other existing normality tests. More precisely we consider the Shapiro-Wilks test, the Anderson-Darling test, the Lilliefors test, the D'Agostino test (see e.g. D'Agostino and Stephens, 1986) and Jarque and Bera test (see Jarque and Bera, 1980).

Références

1. D'Agostino, R.B. and Stephens, M.A. (Eds.) (1986). *Goodness-of-fit techniques*. Statistics: textbooks and monograph, vol. **68**, Marcel Dekker, NY.
2. Goia, A. and Salinelli, E. (2007). Nonlinear Principal Components III. Univariate Distributions, *WP n. 17/2007, Dipartimento SEMeQ, Università del Piemonte Orientale*.
3. Jarque, C.M. and Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, **6**, 255-259.
4. Salinelli, E. (1998). Nonlinear principal components I. Absolutely continuous random variables with positive bounded densities. *Ann. Statist.*, **26 (2)**, 596-616.
5. Salinelli, E. (2008). Nonlinear Principal Components II. Characterization of Normal Distributions. To appear in *Journal of Multivariate Analysis*.

Proximité d'opérateurs unitaires et contiguité des mesures spectrales associées

BOUDOU Alain

Adresse pour correspondance:
 IMT, Université Paul Sabatier, Toulouse
 e-mail: boudou@cict.fr

Résumé

Il est bien connu qu'à tout opérateur unitaire on peut associer, d'une façon biunivoque, une mesure spectrale (à valeurs projecteurs). Ainsi lorsque P est un projecteur orthogonal $U = (I - P) + e^{i\lambda}P$ est un opérateur unitaire dont la mesure spectrale associée est $\delta_0(\cdot)(I - P) + \delta_\lambda(\cdot)P$ (δ_0 et δ_λ étant les mesures de Dirac définies sur \mathcal{B} , tribu de Borel de $[-\pi, \pi[$, respectivement concentrées en 0 et en λ). De même à l'opérateur unitaire I est associé la mesure spectrale $\delta_0(\cdot)I$.

Il est naturel de penser que lorsque deux opérateurs unitaires U et U' sont proches il en est de même, en un certain sens, des mesures spectrales \mathcal{E} et \mathcal{E}' qui leur sont respectivement associées.

Pour mesurer la proximité des opérateurs U et U' nous choisissons la quantité $\sup\{\|U(x) - U'(x)\| ; \|x\| = 1\}$ (égale à $\|U - U'\|_{\mathcal{L}}$). Pour évaluer le voisinage des mesures spectrales \mathcal{E} et \mathcal{E}' on pourrait considérer la quantité $\sup\{\|\mathcal{E}(A) - \mathcal{E}'(A)\|_{\mathcal{L}} ; A \in B\}$. Or un tel choix est inoportun car, par exemple, dans le cas où $U = (I - P) + e^{i\lambda}P$ et $U' = I$, nous obtenons $\|U - U'\|_{\mathcal{L}} = |e^{i\lambda} - 1| \|P\| = |e^{i\lambda} - 1| \leq |\lambda|$ et $\sup\{\|\mathcal{E}(A) - \mathcal{E}'(A)\|_{\mathcal{L}} ; A \in B\} = \sup\{|\delta_0(A) - \delta_\lambda(A)| \|P\|_{\mathcal{L}} ; A \in B\} = 1$. Cela signifie que bien que des opérateurs unitaires puissent être très proche, lorsque λ est petit, la quantité $\sup\{\|\mathcal{E}(A) - \mathcal{E}'(A)\|_{\mathcal{L}} ; A \in B\}$, égale à 1, ne traduit, en aucune façon, un quelconque voisinage des mesures spectrales.

Pour exprimer le voisinage de deux mesures spectrales, nous introduisons et nous étudions la notion d' α -équivalence de mesures spectrales, notion basée sur une relation (bien connue) d'ordre partiel des projecteurs.

On montre alors que la proximité d'opérateurs unitaires, lorsqu'ils commutent, implique celle des mesures spectrales associées au sens de l' α -équivalence. Réciproquement l' α -équivalence de mesures spectrales, lorsqu'elles commutent, implique que les opérateurs unitaires correspondants soient proches. Il est donc clair que la notion d' α -équivalence de mesures spectrales traduit correctement la proximité des opérateurs unitaires.

Nous avons donc un outil qui permet d'appréhender des problèmes de perturbation concernant les opérateurs unitaires. Ceux-ci jouent un grand rôle dans l'étude des processus stationnaires : les opérateurs de "décalage" sont des opérateurs unitaires. Grâce aux résultats précédents nous parvenons à démontrer que si l'opérateur unitaire U associé à une série stationnaire $(x_n)_{n \in \mathbb{Z}}$, c'est-à-dire l'opérateur $x_n \mapsto x_{n+1}$, est tel que U^k est proche de I alors la mesure aléatoire Z dont $(x_n)_{n \in \mathbb{Z}}$ est la transformée de Fourier est concentrée aux voisinages des points λ_q tels que $(e^{i\lambda_q})^k = 1$. Réciproquement si Z est concentrée aux voisinages de tels points du spectre alors l'opérateur de décalage U associé à la série stationnaire $(\int e^{i \cdot n} dZ)_{n \in \mathbb{Z}}$ est tel que U^k est voisin de I . Ce résultat illustre la dualité entre le temporel et le fréquentiel. Nous obtenons pour les fonctions aléatoires stationnaires des résultats analogues.

References

- [1] Arveson, A Short Course on Spectral Theory, Springer, New York, 2001.
- [2] Azencott R., Dacunha-Castelle D., Séries d'observations irrégulières. Modélisation et prévision. Techniques stochastiques, Masson, Paris, 1984.
- [3] Boudou, A, Groupes d'opérateurs unitaires déduit d'une mesures spectrale-une application, C. R. Acad. Sci. Paris, Ser. I 344, 2007.
- [4] Riesz, F., Nagy, B., SZ., Leçons d'analyses fonctionnelle, Gauthier-Villars, Paris, 1968.

Le conflit « Entropie *vs* variance » en analyse de la sensibilité**LANDRI Rémy**

Adresse pour correspondance:
Laboratoire de Probabilités et Statistique
CC51 Université de Montpellier II
Place Eugène Bataillon
34095 Montpellier
e-mail: landri@math.univ-montp2.fr

Dans la réalité, on est souvent amené à prédire, anticiper, contrôler un phénomène engendré par un processus partiellement ou totalement inconnu. Dans ce qui suit, on suppose que la sortie de ce processus est une variable que l'on dénote y^* , dont le comportement est affecté par un groupe de variables $\mathbf{x}^* = (x_1^*, x_2^*, \dots)$. Pour étudier ce comportement, la méthode scientifique nous amène à poser un modèle produisant en sortie une variable à expliquer y ayant quelques prétentions à être proche de la sortie réelle y^* . Ce modèle utilise des variables explicatives x_1, x_2 , etc., que l'on suppose au nombre de m , que l'on regroupe dans le vecteur $\mathbf{x} \equiv (x_1, x_2, \dots, x_m)$. Certains de ces x_i peuvent correspondre à des x_i^* , d'autres non. Ce vecteur est lié à sa valeur de sortie y par une équation mathématique du type $y = f(\mathbf{x})$ complexe.

Dans un contexte environnemental, certains des x_i , ou tous, sont incertains et incorporent donc de la variabilité, c'est-à-dire une incertitude irréductible sur la valeur exacte de l'entrée. Tout devient plus compliqué puisqu'en affectant à ces variables aléatoires (*v.a.*) X_1, X_2 , etc., des distributions de probabilité L_1, L_2 , etc., la sortie Y devient elle-même aléatoire et le degré d'incertitude s'intensifie. Également avec la non-linéarité, la présence d'interactions, ou de fonctions qui s'emboîtent successivement l'une dans l'autre.

On s'interroge ici sur la fiabilité du vecteur \mathbf{X} . Travailler sur ce type de problème, touché jusqu'à présent de façon tangentielle par les statisticiens, présente un intérêt considérable et implique de comprendre quelles sont les variables importantes et donc la sensibilité de chacune dans le modèle.

L'analyse de sensibilité cherche à identifier les incertitudes « importantes » et donc à accroître l'information sur le modèle en quantifiant l'impact des X_i . Elle étudie comment l'incertitude dans la sortie Y d'un modèle peut être attribuée aux différentes sources d'incertitude à l'entrée du modèle [9] en évaluant l'importance de chacune des composantes de \mathbf{X} .

De façon générale, pour mener une analyse de sensibilité, on se focalise sur la variance $\mathbb{V}(Y)$ de la sortie Y . La popularité de la variance, qui exerce pour les modélisateurs une

attraction considérable, provient pour l'essentiel de sa simplicité qui lui impute un rôle central en statistique, à la fois comme une mesure référentielle de la dispersion et comme mesure de la qualité de l'ajustement d'un modèle.

Bien qu'étant pour l'instant assez peu répandue [8], [3] dans le domaine de l'analyse de sensibilité, l'entropie de Shannon,

$$\mathbb{H}(\mathbf{X}) = \left\{ \begin{array}{ll} - \sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) \log p_{\mathbf{X}}(\mathbf{x}) = \mathbb{E}_X [\log p_{\mathbf{X}}(\mathbf{x})] & \text{cas discret} \\ - \int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) \log f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathbb{E}_X [\log f_{\mathbf{X}}(\mathbf{x})] & \text{cas continu} \end{array} \right\},$$

représente une mesure de l'incertitude et de la variabilité qui peut être une alternative à la variance en analyse de la sensibilité. Cette réflexion est née dans le sens où l'entropie donne des indications sur la dispersion des *v.a.*, sans toutefois en apporter autant que la variance, tout comme la variance n'est pas une mesure bien adaptée, et donc critiquable, de l'incertitude. On peut donc chercher à explorer les relations et les différences qu'il existe entre elles.

De ce fait, on examine les rôles de la variance et de l'entropie en ordonnant des distributions multivariées continues en termes de réduction de la mesure. Notamment, l'entropie de Shannon de *v.a.* continues n'est pas invariante lors de transformations inversibles des variables. Aussi, le devenir de l'entropie via des fonctions sera évoqué dans le cas de *v.a.* continues multi-dimensionnelles. On suppose ici que $\mathbf{X} = (X_1, \dots, X_p)$ et $\mathbf{Y} = (X_1, \dots, X_{p-m})$ où $p > m > 1$, et on compare l'effet d'une réduction de la dimension de \mathbf{X} à \mathbf{Y} en termes d'entropie et de variance pour différentes distributions multivariées bien connues. On montre que le concept de réduction d'entropie est plus significatif que celui de réduction de la variance.

On proposera ensuite une comparaison « entropie *vs* variance », à différents niveaux, en termes de variation des paramètres des lois. En ayant une loi univariée $f_X(x)$, la comparaison entropie *vs* variance se ramène à opposer $\mathbb{H}(X)$ *vs* $\mathbb{V}(X)$. Ce problème a d'ailleurs été soulevé par plusieurs auteurs. Maurin [5] s'est intéressé à des comparaisons entre les deux mesures pour les lois gamma et béta. Mukherjee et Ratnaparkhi [6] ont approfondi graphiquement les relations fonctionnelles entre les deux mesures pour différentes lois de probabilité. Ebrahimi, Maasoumi et Soofi [2] s'intéressèrent en particulier au comportement, en termes de croissance ou de décroissance : ce qu'ils appelèrent les ordres de \mathbb{H} et \mathbb{V} , en fonction de variations des paramètres d'échelle, de position ou de forme des distributions. Après avoir discuté du cas univarié, on se focalise sur une variété de familles de distributions bivariées, qui inclue les copules, les lois normales, exponentielles, de Pareto, Gamma, de Dirichlet, et certaines autres lois.

Ce cadre bivarié est beaucoup plus captivant sur les aspects de l'analyse de la sensibilité. Effectivement, en parlant du devenir des entropies et variances de ces lois, on va porter cette recherche sur certains éléments clefs comme $\mathbb{E}_Y \mathbb{V}(\mathbf{X} | \mathbf{Y} = \mathbf{y}^*)$ pour l'analyse de sensibilité basée sur la variance et $\mathbb{H}(\mathbf{X} | \mathbf{Y})$ pour celle basée sur l'entropie.

Finalement, si on considère une loi bivariable $f_{X,Y}(x, y)$, on va donc comparer les différents niveaux de mesures, à savoir :

1. les entropies et les variances marginales :

$$* \mathbb{H}(\mathbf{X}) \text{ vs } \mathbb{V}(\mathbf{X}),$$

$$* \mathbb{H}(\mathbf{Y}) \text{ vs } \mathbb{V}(\mathbf{Y}),$$

2. les entropies des lois conditionnelles et les variances de ces mêmes lois :

$$* \mathbb{H}(\mathbf{X}|\mathbf{Y} = \mathbf{y}^*) \text{ vs } \mathbb{V}(\mathbf{X}|\mathbf{Y} = \mathbf{y}^*),$$

$$* \mathbb{H}(\mathbf{Y}|\mathbf{X} = \mathbf{x}^*) \text{ vs } \mathbb{V}(\mathbf{Y}|\mathbf{X} = \mathbf{x}^*),$$

3. les entropies conditionnelles avec les espérances des variances conditionnelles :

$$* \mathbb{E}_Y \mathbb{H}(\mathbf{X}|\mathbf{Y} = \mathbf{y}^*) \equiv \mathbb{H}(\mathbf{X}|\mathbf{Y}) \text{ vs } \mathbb{E}_Y \mathbb{V}(\mathbf{X}|\mathbf{Y} = \mathbf{y}^*),$$

$$* \mathbb{H}(\mathbf{Y}|\mathbf{X}) \text{ vs } \mathbb{E}_X \mathbb{V}(\mathbf{Y}|\mathbf{X} = \mathbf{x}^*),$$

Et enfin, on doit comparer :

4. l'entropie bivariable $\mathbb{H}(\mathbf{X}, \mathbf{Y})$ vs la matrice des covariances qui s'écrit :

$$\Sigma = \begin{pmatrix} \mathbb{V}(\mathbf{X}) & \text{Cov}(\mathbf{X}, \mathbf{Y}) \\ \text{Cov}(\mathbf{Y}, \mathbf{X}) & \mathbb{V}(\mathbf{Y}) \end{pmatrix}$$

Malheureusement, le problème posé est de savoir sur quelle base comparer $\mathbb{H}(\mathbf{X}, \mathbf{Y})$ avec ce qu'on pourrait envisager comme une « variance bivariable ». En l'état, cette comparaison des deux éléments, l'un scalaire et l'autre une matrice de dimension 2 dans le cas bivariable, est impossible. En revanche, on propose que cette comparaison entre \mathbb{H} et \mathbb{V} ait lieu à partir de certains pivots essentiels de Σ , c'est-à-dire la variance totale : la trace $tr(\Sigma)$, la variance généralisée : le déterminant $|\Sigma|$ et enfin les valeurs propres $\lambda_1(\Sigma)$ et $\lambda_2(\Sigma)$. Ces éléments de Σ permettent cette comparaison avec l'entropie bivariable, qui reste unidimensionnelle.

References

- [1] Darbellay, G.A. et Vajda, I. (2000). Entropy expressions for multivariate continuous distributions. *IEEE Transactions on Information Theory*, **46** : 2, 709–712.

- [2] Ebrahimi, N., Maasoumi, E. et Soofi, E.S. (1999). Ordering univariate distributions by entropy and variance. *Journal of Econometrics*, **90**, 317–336.
- [3] Krzykacz-Hausmann, B. (2001). Epistemic sensitivity analysis based on the concept of entropy. *Proceedings of SAMO 2001*, eds. P. Prado and R. Bolado, Madrid, CIEMAT, 31–35.
- [4] Landri, R., Ducharme, G. et Bonnard, R. (2007). Entropy-based sensitivity analysis in health risk assessment. *Fifth International Conference on Sensitivity Analysis of Model Output*, Eötös University, Budapest, Hungary, 18–22 June.
- [5] Maurin, M. (1999). Jeux de variance et d'entropie. *XXXI Journées de Statistique de la SFDS*, 17-21 mai, Grenoble, France, 4p.
- [6] Mukherjee, D., Ratnaparkhi, M.V. (1986). On the functional relationship between entropy and variance with related applications. *Communications in Statistics-Theory and Methods*, **15**, 291–311.
- [7] Nadarajah, S., Zografos, K. (2005). Expressions for Rényi and Shannon entropies for bivariate distributions. *Informations Sciences*, **170**, 173–189.
- [8] Sacks, J., Welch, W.J., Mitchell, T.J. et Wynn, H.P. (1989). Design and Analysis of Computer Experiments. *Statistical Science*, **4** : 4, 409–435.
- [9] Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis*, **22** : 3, 579–590.
- [10] Shannon, C.E. (1948). A mathematical theory of communication. *Bell system technical journal*, **27**, 379–423 (I-II) and 623–656 (III-IV).

k-Nearest Neighbor Method in functional nonparametric regression

BURBA Florent *

* Adresse pour correspondance:
I.M.T. UMR 5219 Université Paul Sabatier
Toulouse, France
e-mail: burba@cict.fr

Résumé

During the last few years, the use of functional datasets (i.e. one observation contains at least a functional object like curve, surface or image...) became very usual. Thanks to technological progress, measurement techniques are now more powerful and the quantity of collected informations becomes voluminous with a great precision. In fact, instants of measure come closer so grids are more and more fine. These practical improvements have naturally led to the development of the theory of Functional Statistics. Functional data analysis has been popularized by Ramsay & Silverman (1997, 2002, 2005). Considering the high-dimensionnal aspect of the data, it is natural to use nonparametric models and theoretical tools for estimation, classification and discrimination has been studied by Ferraty and Vieu (2006). In particular, they presented the functional adaptation of Nadaraya-Watson type kernel regression estimator. One of the most important concern in high-dimensional problems is to take into account the local structure of the data and it appears that a well adapted tool is the k nearest neighbor (kNN) method.

kNN method for estimating regression has been widely used and studied for many years in the multivariate case. The first works were presented by Royall (1966) and Stone (1977). Mack (1981) studied the L^2 convergence and the asymptotic distribution, and Devroye (1978, 1981) presented the strong consistency and the uniform convergence. For complete studies, see, for examples, Bosq and Lecoutre (1987) or Györfi *et al* (2002).

The real interest of kNN method comes from the nature of the smoothing parameter. Indeed, in the traditional kernel method, the smoothing parameter is the bandwidth h , which is a real positive number. Here, the number of neighbors k is the smoothing parameter

and it takes its values in a discrete set. As we said previously, the other very important aspect of this method is that it allows the construction of a neighborhood adapted to the local structure of the data. The main difficulties with kernel method appear when data are sparse, choosing the number of neighbors allows to avoid this problem and is adapted to the concentration of the data. The importance of local adaptive bandwidth is well-known in nonparametric statistics and this is even more true with functional data because of the high complexity of infinite dimensional spaces (see Benhenni *et al* (2007) and Fernandez de Castro *et al* (2005) for instance). We will see here how kNN method provides a simple and efficient way to build local bandwidth.

However, kNN method presents a major technical difficulty: the selection of the nearest neighbors gives a random bandwidth. The second problem, linked with the functional nature of the data, is that we don't suppose the existence of a density (because, in infinite dimensional spaces, we don't have any standard measure as the Lebesgue measure in the multivariate case). We will see here a first theoretical study of this method for the estimation of nonparametric regression in the functional setting.

This talk will be organised as follows: we will first present the model and the kNN estimator. Then, we will give asymptotic properties of the method: the pointwise almost complete convergence and his rate. Finally, we will illustrate the effectiveness of kNN method by comparing it with the standard kernel approach on two examples. In the first one, we will use a finite size simulated curves datasets and, in the second one, we will study spectrometric curves. These examples emphasize that kNN method is more and more efficient when the concentration of the curves data presents more and more heterogeneity (we will also explain and illustrate what is heterogeneity in our situation).

All results presented in this talk stem from Burba *et al* (2008a and 2008b).

Références

K. Benhenni et al. (2007), *Local smoothing regression with functional data*, Comput. Statist. 22, pp 353–369.

D. Bosq and J.P. Lecoutre (1987), *Théorie de l'estimation fonctionnelle*, Economica.

Burba F., Ferraty F., Vieu P. (2008a), *Convergence de l'estimateur à noyau des k plus proches voisins en régression fonctionnelle non-paramétrique*, C. R. Acad. Sci. Paris, Ser. I 346.

Burba F., Ferraty F., Vieu P. (2008b), *k-Nearest Neighbor Method in functional nonparametric regression*, submitted to Journal of Nonparametric Statistics.

L.P. Devroye (1978), *The uniform convergence of nearest neighbor regression function estimators and their application in optimization*, IEEE. Trans. Inform. Theory 24, pp 142–151.

L.P. Devroye (1981), *On the almost everywhere convergence of nonparametric regression function estimates*, Ann. Statist. 9, pp 1310–1319.

B. Fernandez de Castro, S. Guillas and W. Gonzalez Manteiga (2005), *Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels*, Technometrics 47, pp 212-222.

F. Ferraty and P. Vieu (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer, New York.

L. Györfi et al. (2002), *A Distribution-Free Theory of Nonparametric Regression*, Springer Series in Statistics, Springer, New York.

Mack Y.P. (1981), *Local properties of kNN regression estimates*, J. Alg. Disc. Meth. 2, pp 311–323.

J. Ramsay and B.W. Silverman (1997), *Functional Data Analysis*, Springer-Verlag, New York.

J. Ramsay and B.W. Silverman (2002), *Applied Functional Data Analysis: Methods and case studies*, Springer-Verlag, New York.

J. Ramsay and B.W. Silverman (2005), *Functional Data Analysis. 2nd edition*, Springer-Verlag, New York.

R.M. Royall (1966), *A class of nonparametric estimates of a smooth regression function*, Ph.D. diss., Stanford University.

C.J. Stone (1977), *Consistent nonparametric regression*, Ann. Statist. 5, pp 595–645.

Estimation semi-paramétrique de distribution de dénombrement

SENGA KIESSE Tristan

Adresse pour correspondance:

Université de Pau et des Pays de l'Adour
Laboratoire de Mathématiques Appliquées - UMR 5142 CNRS
Avenue de l'Université - 64000 Pau, France.
E-mail : tristan.sengakiesse@univ-pau.fr

Dans cet exposé, nous proposons un estimateur semi-paramétrique de distribution de dénombrement sous l'hypothèse de modèle de Poisson pondéré ou binomial pondéré. L'estimation non-paramétrique de la fonction discrète de poids correspondant au modèle est effectuée en utilisant la méthode des noyaux associés discrets. Des modèles de diagnostic sont mis en place pour aider au choix d'une approche purement paramétrique, semi-paramétrique ou non-paramétrique.

Bayesian Collocation Tempering for Nonlinear Differential Equation Models

CAMPBELL Dave* and STEELE Russell

Adresse pour correspondance:
Simon Fraser University, 8888 University drive
Burnaby, BC, Canada V5A1S6
e-mail: dac5@sfu.ca

Résumé

When functional observations are thought to arise from noisy measurements $y(t)$ of a system of a nonlinear differential equation (DE) models based on parameters θ :

$$\begin{aligned} \frac{dx(t)}{dt} &= f(x, \theta) \\ y(t) &\sim N(x(t), \sigma^2), \end{aligned}$$

fitting the model to the data produces unique challenges (Gelman, Bois, Jiang 1996). Since closed form solutions to DEs are rarely available, the likelihood and posterior do not have a closed form. Numerical methods can be used to produce a DE solution, however this requires estimating the initial states $x(0)$, augmenting the dimension of the unknown parameter vector. Furthermore, the variety of dynamics that a DE model can describe may produce posterior topologies with multiple minor modes, ridges, ripples and flat sections (Deuffhard, Bornemann 2000). Consequently, when the likelihood is centered on the numerical solution to the system of DEs gradient based methods and standard MCMC can converge to minor modes associated with partial fits to the data.

We propose a new method, Collocation Tempering, a population based MCMC method (see Jasra, Stephens, and Holmes (2007) for a recent overview) where parallel MCMC chains use approximations to the DE model based on parametric smoothing. Each of the parallel MCMC chains uses a likelihood centered on a smooth, balancing the fit to the data with the dynamics of the DE model, thereby allowing some deviation from the numerical DE solution. The resulting model benefits from the simultaneous data feature matching and DE model matching capabilities of model based smoothing. This marriage of parametric model based

smoothing and population MCMC does not depend on the numerical DE solution or require estimates of $\hat{x}(0)$. Furthermore this methods substantially improves convergence times and robustness to prior distributions that disagree with information in the data and values used to initialize the algorithm while requiring fewer parallel chains compared to other population based MCMC methods.

Références

Deuffhard, P. and Bornemann, F. (2000). *Scientific Computing with Ordinary Differential Equations*. Springer.

Gelman, A., Bois, F. Y. and Jiang, J. (1996). *Physiological pharmacokinetic analysis using population modeling and informative prior distributions*. Journal of the American Statistical Association 91 (436), 1400-1412

Jasra, A., Stephens, D. A., and Holmes, C. C. (2007), *On Population-Based Simulation for Static Inference*, Statistics and Computing, 17, 263-279.

Modélisation fonctionnelle de courbes de densités

MORLAIS Fabrice

Adresse pour correspondance:
ERI3 INSERM "Cancers & Populations"
EA 3936 Université Caen
Unité de Recherche et d'Evaluation en Epidémiologie
Pôle de Santé des Populations
Faculté de médecine - avenue côte de nacre - 14032 Caen cedex
e-mail: fabricemorlais@yahoo.fr

Résumé

En modélisation fonctionnelle, l'individu n'est plus seulement représenté par un ensemble de variables mais par une fonction (à valeur dans un espace infini). Une variable aléatoire χ est considérée comme fonctionnelle si elle prend des valeurs dans un espace infini. Par exemple la variable fonctionnelle $\chi = \{X(t); t \in T\}$ avec $T \in \mathfrak{R}$ représentera une courbe observée sur l'intervalle T de \mathfrak{R} .

La première étape à considérer dans une modélisation fonctionnelle est la transformation des données initialement discrétisées en une fonction continue. Lorsque la fonction à estimer est une densité de probabilité de nombreuses techniques statistiques existent pour réaliser cette transformation, parmi les plus populaires citons la méthode non paramétrique d'estimateur à noyau préconisée par Sheather et Jones et la méthode par splines basiques avec contrainte de Gu.

Bien que les méthodes statistiques classiques soient à peu près toutes développées dans le cadre fonctionnel, quelques difficultés surviennent pour les adapter lorsque les données fonctionnelles considérées sont des densités. Les méthodes statistiques listées ci-dessous sont des exemples permettant d'utiliser la densité de probabilité comme unité statistique.

En statistique exploratoire, Kneip a développé une ACP fonctionnelle adaptée aux densités. Celle-ci s'intéresse, par l'intermédiaire d'un estimateur à noyau, à la matrice des corrélations des fonctions de densité, pour réaliser l'analyse exploratoire.

En statistique supervisée, la méthode non paramétrique développée par Ferraty et Vieu s'adapte aux densités puisqu'elle repose essentiellement sur la notion de distance entre courbes. Alors que la notion de distance est facile à définir dans un espace fini, celle-ci est beaucoup plus difficile à définir dans un espace infini. L'utilisation de semi-métrie permet de contourner ce problème.

Pour comparer globalement des courbes de densité, Delicado a développé une méthode modifiant l'ANOVA fonctionnelle de Cuevas pour la rendre adaptable aux fonctions de densité de probabilité. Cette méthode utilise, là encore, la notion de distance entre courbes et les tests de permutation.

Références

1. Silverman BW. Density Estimation for Statistics and Data Analysis. 1986.
2. Adrian W, Bowman, Adelchi Azzalini, Applied Smoothing Techniques for Data Analysis. 1997.
3. Wasserman L. All of Nonparametric Statistics. 2007.
4. Härdle. Applied Nonparametric Regression. 1992.
5. Gu C. Smoothing Spline ANOVA models. 2002.
6. Ramsay JO, Silverman BW. Applied Functional Data Analysis Methods and Case Studies. 2002.
7. Ramsay JO, Silverman BW. Functional Data Analysis Second Edition. 2005.
8. Ferraty F, Vieu P. Non Parametric Functional Data Analysis Theory and Practice. 2006.
9. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density method.
10. Kneip A. Inference for density families using functional principal component analysis.
11. Delicado P. Functional k-sample problem when data are density functions.