Density-based inverse calibration with functional predictors

Nathalie Villa-Vialaneix

http://www.nathalievilla.org

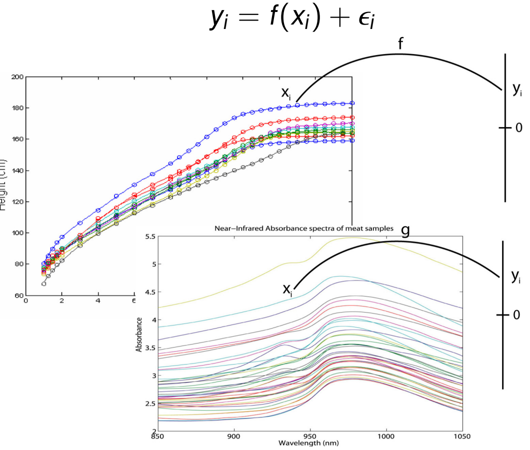
nathalie.villa@toulouse.inra.fr

ICOR 2014, Havana, March, 12th

Joint work with **Noslen Hernández**, **Rolando J. Biscay** and **Isneri Talavera**

Introduction

The fast development of instrumental analysis equipment provides huge amount of data as **high-resolution digitized functions**.



Data

generally represented by **high-dimensional vectors** whose values at different coordinates are **strongly correlated**.

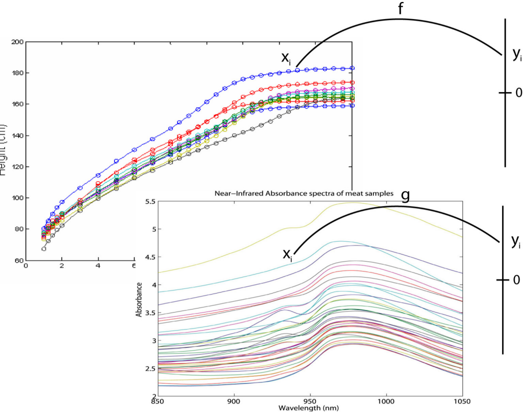
Curse of dimensionality

Usually the dimension of such vectors greatly exceeds the sample size.

Introduction

The fast development of instrumental analysis equipment provides huge amount of data as **high-resolution digitized functions**.

$$y_i = f(x_i) + \epsilon_i$$



ill-posed problem

The direct application of classical multivariate regression methods for this type of data often leads to the inversion of an ill-conditioned (variance) matrix

Functional Data Analysis

Functional Data Analysis (FDA): statistical techniques that take into account the **functional** nature of the data.

- first developed by **[Ramsay and Silverman, 1997]**
- **Idea**: think about observed data as **one or more continuous real-valued functions**, rather than as finite-dimensional vectors.
- also integrate some **functional processing techniques** (derivation, integration, etc).

Example of FDA methods: $X \in L_2([0,1])$, $Y \in \mathbb{R}$

Purpose: approximate the regression function: $\gamma(x) = \mathbb{E}(Y|X = x)$

a) Functional linear methods

- [Cardot et al., 1999]

b) Functional nonparametric regression methods

- Functional Kernel (NWK) [Ferraty and Vieu, 2006]
- Functional Neural Networks (NN, NN-RBF) [Rossi et al., 2005]
- Functional regression in RKHS (RBF, FSVR) [Preda, 2007, Hernández et al., 2007]

c) Functional inverse regression

- Functional sliced inverse regression (FIR) [Ferré and Yao, 2003, Ferré and Yao, 2005]

Outline

1. Functional Density-Based Inverse Regression (DBIR)
2. Simulations
3. Concluding Remarks

Definition of DBIR

Data:

(X, Y)	random variables taking values in $\mathcal{X} \times \mathbb{R}$
$\mathcal{D} = (x_1, y_1), ..., (x_n, y_n)$	independent realizations of (X, Y)

Definition of DBIR

Data:

(X, Y)	random variables taking values in $\mathcal{X} \times \mathbb{R}$
$\mathcal{D} = (x_1, y_1), ..., (x_n, y_n)$	independent realizations of (X, Y)

To be estimated...

The regression function

$$\gamma(X) = \mathbb{E}(Y|X = x)$$

Definition of DBIR

Data:

(X, Y)	random variables taking values in $\mathcal{X} \times \mathbb{R}$
$\mathcal{D} = (x_1, y_1), \ldots, (x_n, y_n)$	independent realizations of (X, Y)

To be estimated...

The regression function

$$\gamma(X) = \mathbb{E}(Y|X = x)$$

The "true" physical model is

$$X = F(Y) + \epsilon, \qquad F : \mathbb{R} \to \mathcal{X}$$

where ϵ: stochastic process with zero mean, independent of Y.

Assumptions for DBIR

1. There exists a probability measure P_0 on $\mathcal{X}$ (not depending on Y) such that $P(.|y)$ is absolutely continuous with respect to P_0 ($P \ll P_0$).

$$\Downarrow$$

$\exists\, f(.|y) = \frac{dP(A|y)}{dP_0}$ such that $P(A|y) = \int_A f(.|y)dP_0$, $\forall\, A \subset \mathcal{X}$ measurable.

2. Y is a continuous random variable.

$$\Downarrow$$

$\exists\, f_Y(y)$

Proposal

$$\begin{aligned}\gamma(x) &= \mathbb{E}(Y|X=x)\\ &= \int_{\mathbb{R}} y f(y|x) dy\\ &= \int_{\mathbb{R}} y \frac{f(x|y) f_Y(y)}{f_X(x)} dy\\ &= \boxed{\frac{1}{f_X(x)} \int_{\mathbb{R}} y f(x|y) f_Y(y) dy}\end{aligned}$$

$$f(y|x) = \frac{f(x|y) f_Y(y)}{f_X(x)}$$

$$f_X(x) = \int_{\mathbb{R}} f(x|y) f_Y(y) dy$$

Proposal

$$\begin{aligned}\gamma(x) &= \mathbb{E}(Y|X=x)\\ &= \int_{\mathbb{R}} y f(y|x) dy \\ &= \int_{\mathbb{R}} y \frac{f(x|y) f_Y(y)}{f_X(x)} dy \\ &= \boxed{\frac{1}{f_X(x)} \int_{\mathbb{R}} y f(x|y) f_Y(y) dy}\end{aligned}$$

$$f(y|x) = \frac{f(x|y) f_Y(y)}{f_X(x)}$$

$$f_X(x) = \int_{\mathbb{R}} f(x|y) f_Y(y) dy$$

General estimator

$$\hat{\gamma}(x) = \frac{\sum_{i=1}^n \hat{f}(x|y_i) y_i}{\sum_{i=1}^n \hat{f}(x|y_i)}$$

Gaussian case

Additional assumption with $\mathcal{X} = L_2([0, 1])$

- For a fixed y, $P(.|y)$ is a Gaussian measure on $\mathcal{X}$ with mean $\mu(y)$ and covariance operator Γ.

Gaussian case

Additional assumption with $\mathcal{X} = L_2([0,1])$

- For a fixed y, $P(.|y)$ is a Gaussian measure on $\mathcal{X}$ with mean $\mu(y)$ and covariance operator Γ.

Consequences

for $(\varphi_j, \lambda_j)_j$ eigenfunctions and eigenvalues of Γ:

- $$\sum_{j=1}^{\infty} \frac{\mu_j^2(y)}{\lambda_j} < \infty, \qquad \mu_j(y) = \langle \mu(y), \varphi_j \rangle$$
- $$f(x|y) = \exp\left\{ \sum_{j=1}^{\infty} \frac{\mu_j(y)}{\lambda_j} \left(x_j - \frac{\mu_j(y)}{2} \right) \right\}, \qquad x_j = \langle x, \varphi_j \rangle, \ \forall j \geq 1$$

Specification in the Gaussian case

1. Obtain an estimate $\hat{\mu}(y)$ of $\mu(y)$

Nadaraya-Watson kernel estimate for $\mu(y,t) = \mathbb{E}(X(t)|Y=y)$

$$\hat{\mu}(y,t) = \frac{\sum_{i=1}^n x_i(t) K\left(\frac{y_i - y}{h}\right)}{\sum_{i=1}^n K\left(\frac{y_i - y}{h}\right)}$$

Specification in the Gaussian case

1. Obtain an estimate $\hat{\mu}(y)$ of $\mu(y)$

Nadaraya-Watson kernel estimate for $\mu(y,t) = \mathbb{E}(X(t)|Y = y)$

$$\hat{\mu}(y,t) = \frac{\sum_{i=1}^n x_i(t) K\left(\frac{y_i - y}{h}\right)}{\sum_{i=1}^n K\left(\frac{y_i - y}{h}\right)}$$

2. Obtain estimates $(\hat{\lambda}_i, \hat{\varphi}_i)_i$ of $(\lambda_i, \varphi_i)_i$ on the basis of the empirical covariance of the residuals

Functional PCA

$$\hat{\epsilon}_i = x_i - \hat{\mu}(y_i), \qquad i = 1, ..., n$$

$$\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^n (\hat{\epsilon}_i - \overline{\epsilon}) \otimes (\hat{\epsilon}_i - \overline{\epsilon}), \qquad \overline{\epsilon} = \frac{1}{n} \sum_{i=1}^n \epsilon_i$$

Specification in the Gaussian case II

3. Chose $p = p(n) \in \mathbb{N}$,

Estimate $f(x|y)$

$$\hat{f}(x|y) = \exp\left\{\sum_{j=1}^{p} \frac{\hat{\mu}_j(y)}{\hat{\lambda}_j}\left(\hat{x}_j - \frac{\hat{\mu}_j(y)}{2}\right)\right\}, \text{ with } \left\{\begin{array}{l} \hat{\mu}_j(y) = \langle \hat{\mu}(y), \hat{\varphi}_j \rangle \\ \text{and } x_j = \langle x, \hat{\varphi}_j \rangle \end{array}\right.$$

Specification in the Gaussian case II

3. Chose $p = p(n) \in \mathbb{N}$,

Estimate $f(x|y)$

$$\hat{f}(x|y) = \exp\left\{\sum_{j=1}^{p} \frac{\hat{\mu}_j(y)}{\hat{\lambda}_j}\left(\hat{x}_j - \frac{\hat{\mu}_j(y)}{2}\right)\right\}, \text{ with } \left\{\begin{array}{l} \hat{\mu}_j(y) = \langle \hat{\mu}(y), \hat{\varphi}_j \rangle \\ \text{and } x_j = \langle x, \hat{\varphi}_j \rangle \end{array}\right.$$

DBIR estimator

$$\hat{\gamma}(x) = \frac{\sum_{i=1}^{n} \hat{f}(x|y_i) y_i}{\sum_{i=1}^{n} \hat{f}(x|y_i)}$$

Consistency DBIR

Theorem [Hernández et al., 2014]

Under assumptions (A1)-(A12), we have, for all $x \in \mathcal{X}$ such that $f_X(x) > 0$,

$$\lim_{n \to +\infty} \hat{\gamma}(x) =^{\mathbb{P}} \gamma(x).$$

Outline

Simulations

- (X, Y): random pair taking values in $\mathcal{X} = L_2([0,1]) \times \mathbb{R}$
- $(v_j)_{j\geq 1}$: the trigonometric basis of $\mathcal{X}$:

$$\begin{aligned} v_{2k-1} &= \sqrt{2}\cos(2\pi k) \\ v_{2k} &= \sqrt{2}\sin(2\pi k) \quad \text{for} \quad k = 1, 2, \ldots \end{aligned}$$

- $X = \mu(Y) + \epsilon$ with ϵ: Gaussian process (independent of Y) with zero mean and covariance operator

$$\Gamma_\epsilon = \sum_{j\geq 1} \frac{1}{j} v_j \otimes v_j$$

Simulated models

M2

$$X = Yv_1 + 2Yv_2 + 3Yv_{10} + \epsilon$$

M2

$$X = \sin(Y)v_1 + \log(Y + 1)v_5 + \epsilon$$

M3

$$X = Yq_1 + 5Yq_2 + \epsilon$$

where $q_1 = 2t^3$ and $q_2 = t^4$

M4

$$X = \sin(Y)q_1 + 20\log(Y + 1)q_2 + \epsilon$$

Computational aspects of the simulations

- $n_L = 300$ (number of samples in the training)
- $n_T = 200$ (number of samples in the test)
- $t \in \tau$, where τ equally spaced into $[0, 1]$
- $Y \sim \mathcal{U}(0, 10)$;
- ϵ was simulated by using a truncation of Γ_ϵ:

$$\Gamma_\epsilon(s, t) \simeq \sum_{j=1}^{N_j} \frac{1}{j} v_j(t) v_j(s)$$

with $N_j = 500$.

Performances compared to standard nonparametric estimate

Model	DBIR	NWK	DBIR (linear est. of the mean)
$M1$	0.23	0.28	0.22
$M2$	1.71	1.91	X
$M3$	0.07	0.19	0.02
$M4$	0.35	0.47	X

Outline

Conclusion and Remarks

- DBIR based on **estimation of the "inverse" regression model** $f(x|y)$ but (contrary to FSIR) no need to choose the effective dimension d and only 1-dimensional regression (instead of d);
- **computationally simple** and performs well on the simulated data;
- $P(.|y)$ assumed Gaussian but no assumption on the distribution of X or Y or Y given X.

Thank you for your attention...

...any question?

References

Cardot, H., Ferraty, F., and Sarda, P. (1999).
Functional linear model.
Statistics and Probability Letters, 45:11–22.

Ferraty, F. and Vieu, P. (2006).
NonParametric Functional Data Analysis.
Springer.

Ferré, L. and Yao, A. (2003).
Functional sliced inverse regression analysis.
Statistics, 37(6):475–488.

Ferré, L. and Yao, A. (2005).
Smoothed functional inverse regression.
Statistica Sinica, 15(3):665–683.

Hernández, N., Biscay, R., and Talavera, I. (2007).
Progress in Pattern Recognition, Image Analysis and Applications (Proceedings of 12th Iberoamericann Congress on Pattern Recognition, CIARP 2007, Valparaiso, Chile, November 13-16, 2007), volume 4756 of *Lecture Notes in Computer Science*, chapter Support vector regression methods for functional data, pages 564–573.
Springer, Berlin/Heidelberg.

Hernández, N., Biscay, R., Villa-Vialaneix, N., and Talavera, I. (2014).
A non parametric approach for calibration with functional data.
In revision for publication in Statistica Sinica.

Preda, C. (2007).
Regression models for functional data by reproducing kernel Hilbert spaces methods.
Journal of Statistical Planning and Inference, 137(3):829–840.

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

Rossi, F., Delannay, N., Conan-Guez, B., and Verleysen, M. (2005).
Representation of functional data in neural networks.
Neurocomputing, 64:183–210.

Consistency of the DBIR estimator (notations)

- m st $\mu(.|y) = \frac{m(y)}{f_Y(y)}$;
- $(\eta_n)_n \subset \mathbb{R}^{+*}$, $\eta_n \searrow 0$;
- $r : y \rightarrow \mu(.|y)$, $r_{\eta_n}(y) = \frac{m(y)}{f_{Y,\eta_n}(y)}$ with $f_{Y,\eta_n}(y) = \max(f_Y(y), \eta_n)$;
- $\hat{m}(y) = \sum_{i=1}^n x_i K\left(\frac{y_i - y}{h}\right)$;
- $\hat{r}_{\eta_n} : y \rightarrow \hat{\mu}(.|y), \quad \hat{r}_{\eta_n}(y) = \frac{\hat{m}(y)}{\hat{f}_{Y,\eta_n}(y)}, \quad \hat{f}_{Y,\eta_n}(y) = \max(\hat{f}_Y(y), \eta_n)$;
- $\hat{f}_Y(y) = \sum_{i=1}^n K\left(\frac{y_i - y}{h}\right)$.

Consistency of the DBIR estimator (part 1)

(A1) f_Y has support $\Omega_Y \subset \mathbb{R}$, and f_Y and r are $\mathcal{C}^k$, for a $k \geq 2$, on Ω_Y;

(A2) K is an order k kernel with compact support;

(A3) there exists d_1 and d_2 such that $\sup_{y\in\Omega_Y} \left|f_Y^{(k)}(y)\right| < d_1$ and $\sup_{y\in\Omega_Y} \left\|r^{(k)}(y)\right\| < d_2$;

(A4) $h = O\left(n^{-c_1}\right)$, where $\frac{1}{4+2k} < c_1 < \frac{1}{4}$;

(A5) there exists $b_1 > 0$ such that $\inf_{y\in\Omega_Y} f_Y(y) \geq b_1$;

(A6) there exists $b_2 > 0$ such that $\sup_{y\in\Omega_Y} \left\|r(y)\right\| \leq b_2$.

Proposition 1

Under assumptions (A1)-(A6) we have:
$\sup_{y\in\Omega_Y} \left\|\hat{r}(y) - r(y)\right\| = O_P\left(n^{-c_1 k} + \left(\frac{\log n}{n^{1-2c_1}}\right)^{1/2}\right)$.

Consistency of the DBIR estimator (part 2)

(A7) $\mathbb{E}\left(\|\epsilon\|^4\right) < +\infty$;

Proposition 2

Under assumptions (A1)-(A7) we have:

$$\left\|\Gamma - \hat{\Gamma}\right\| = O_P\left(\frac{1}{n^{1/2-2c_1}}\right).$$

Consistency of the DBIR estimator (part 3)

(A8) $\sum_{j=1}^{\infty} \sup_{y\in\Omega_Y} \frac{|r_j(y)|}{\sqrt{\lambda_j}} < \infty$;

(A9) The $(\lambda_j)_j$ are all distinct;

(A10) $\lim_{n\to+\infty} p = +\infty$;

(A11) $\lim_{n\to+\infty} \frac{\sum_{j=1}^{p} a_j}{\lambda_p n^{1/2-2c_1}} = 0$;

(A12) $\frac{p}{\lambda_p^2} = O\left(n^q\right)$ for some $0 < q < \min(c_1 k, \frac{1}{2} - c_1)$.

Proposition 3

Under Assumptions (A1)-(A12), for any $x \in \mathcal{X}$ we have:
$\sup_{y\in\Omega_Y} \left|\hat{f}(x|y) - f(x|y)\right| = o_P\left(1\right)$.