A functional density-based nonparametric approach for statistical calibration

No Author Given

No Institute Given

Abstract. In this paper a new nonparametric functional method is introduced for predicting a scalar random variable Y from a functional random variable X. The resulting prediction has the form of a weighted average of the training data set, where the weights are determined by the conditional probability density of X given Y, which is assumed to be Gaussian. In this way such a conditional probability density is incorporated as a key information into the estimator. Contrary to some previous approaches, no assumption about the dimensionality of $\mathbb{E}(X|Y = y)$ is required. The new proposal is computationally simple and easy to implement. Its performance is shown through its application to both simulated and real data.

1 Introduction

The fast development of instrumental analysis equipment and modern measurement devises provides huge amounts of data as high-resolution digitized functions. As a consequence, Functional Data Analysis (FDA) has become a growing research field. In the FDA setting, each individual is treated as a single entity described by a continuous real-valued functions rather than by a finite-dimensional vector: functional data (FD) are then supposed to have values in an infinite dimensional space, often particularized as a Hilbert space.

An extensive review of the methods developed for FD can be found in the monograph of Ramsay and Silverman [1]. In the case of functional regression, where one intends to estimate a random scalar variable Y from a functional variable X taking values in a functional space $\mathcal{X}$, earlier works were focused on linear methods such as the functional linear model with scalar response [2–8] or the functional Partial Least Squares [9]. More recently, the problem has also been addressed nonparametrically with smoothing kernel estimates [10], multilayer perceptrons [11], and support vector regression [12, 13]. Another point of view between these two approaches is to use a semi-parametric approach, such as the SIR (Sliced Inverse Regression, [14]) that has been extended to functional data in [15–17]. In this approach, the functional regression problem is addressed

through the opposite regression problem i.e., the estimation of $\mathbb{E}(X|Y=y)$, by assuming that this quantity belongs a finite dimensional subspace of $\mathcal{X}$.

In this paper, a similar approach is presented: we rely on the estimation of the regression of X on Y to estimate the regression function $\gamma(X) = \mathbb{E}(Y|X)$ but, contrary to the SIR approach, no assumption on the dimensionality of $\mathbb{E}(X|Y=y)$ is required, and furthermore the specific form of the conditional probability density of X given Y, which is assumed to be Gaussian, is incorporated as a key information into the estimator. A practical motivation to the latter model arises from calibration problems in chemometrics, specifically in spectroscopy, where some chemical variable (e.g., concentration) needs to be predicted from a digitized function (e.g., spectra). In this setting, the spectral function is the output of a physical data generation process in which the scalar variable of interest (i.e, concentration) is the input, plus some random perturbation due to the measurement procedure. The introduced method, which will be referred to as functional Density-Based Nonparametric Regression (DBNR), is computationally simple and easy to implement.

This paper is divided as follows. Section 2 presents the functional Density-Based Nonparametric Regression method. Then, Sections 3 and 4 illustrate the use of this approach in a simulation study and in a real-world problem coming from chemometrics. Conclusions are given in Section 5.

2 Functional Density-Based Nonparametric Regression

2.1 Definition of DBNR in a general setting

Let (X, Y) be a pair of random variables taking values in $\mathcal{X} \times \mathbb{R}$ where $(\mathcal{X}, \langle ., . \rangle)$ is a Hilbert space. Suppose also that n i.i.d. realizations of (X, Y) are given, denoted by $(x_i, y_i)_{i=1,\ldots,n}$. The goal is to build, from $(x_i, y_i)_i$, a way to predict a new value for Y from a given (observed) value of X. This problem is usually addressed by the estimation of the regression function $\gamma(x) = \mathbb{E}(Y|X=x)$.

The functional density-based nonparametric regression implicitly supposes that the inverse model makes sense; this inverse model is:

$$X = F(Y) + \epsilon \tag{1}$$

where ϵ is a random process (perturbation or noise) with zero mean, independent of Y, and $y \to F(y)$ is a function from $\mathbb{R}$ into $\mathcal{X}$. As was stated in Section 1, this is a common background for calibration problems, amongs others.

Additionally, the following assumptions are made: first, it exists a probability measure P_0 on $\mathcal{X}$ (not depending on y) such that the conditional probability measure of X given $Y = y$, say $P(\cdot / y)$, has a density $f(\cdot / y)$ with respect to P_0:

$$P(A / y) = \int_A f(x / y) \, P_0(dx)$$

for any measurable set A in $\mathcal{X}$. Furthermore, it is assumed that Y is a continuous random variable, i.e., that its distribution has a density $f_Y(y)$ (with respect to the Lebesgue measure on $\mathbb{R}$).

Under these assumptions, the regression function is:

$$\gamma(x) = \frac{\int_{\mathbb{R}} f(x/y)\, f_Y(y)\, y dy}{f_X(x)}, \quad \text{where} \quad f_X(x) = \int_{\mathbb{R}} f(x/y)\, f_Y(y)\, dy.$$

Hence, given an estimate $\widehat{f}(x/y)$ of $f(x/y)$, the following estimate of $\gamma(x)$ can be constructed from the previous equation:

$$\widehat{\gamma}(x) = \frac{\sum_{i=1}^{n} \widehat{f}(x/y_i)\, y_i}{\widehat{f}_X(x)}, \quad \text{where} \quad \widehat{f}_X(x) = \sum_{i=1}^{n} \widehat{f}(x/y_i). \tag{2}$$

2.2 Specification in the Gaussian case

The general estimation scheme given in Equation (2) will be here specified for the case in which $P(\cdot/y)$ is a Gaussian measure on $\mathcal{X} = \mathcal{L}_2[0,1]$ for each $y \in \mathbb{R}$. $P(\cdot/y)$ is then supposed to have a mean function $\mu(\cdot/y) \in \mathcal{X}$ (which is then equal to $F(y)(\cdot)$ according to Equation (1)) and a covariance operator r (not depending on y), which is a Hilbert-Schmidt operator on the space $\mathcal{X}$. Then, there exists an eigenvalue decomposition of r, $(\varphi_j, \lambda_j)_{j\geq 1}$ such that $(\lambda_j)_j$ is a decreasing series of positive real numbers, $(\varphi_j)_j$ take values in $\mathcal{X}$ and $r = \sum_j \lambda_j \varphi_j \otimes \varphi_j$ where $\varphi_j \otimes \varphi_j(h) = \langle \varphi_j, h\rangle \varphi_j$ for any $h \in \mathcal{X}$.

Denote by P_0 the Gaussian measure on $\mathcal{X}$ with zero mean and covariance operator r. Assume the following usual regularity condition holds: for each $y \in \mathbb{R}$,

$$\sum_{j=1}^{\infty} \frac{\mu_j^2(y)}{\lambda_j} < \infty, \qquad \text{with } \mu_j(y) = \langle \mu(\cdot/y), \varphi_j \rangle.$$

Then, $P(\cdot/y)$ and P_0 are equivalent Gaussian measures, and the density $f(\cdot/y)$ has the explicit form:

$$f(x/y) = \exp\left\{ \sum_{j=1}^{\infty} \frac{\mu_j(y)}{\lambda_j} \left(x_j - \frac{\mu_j(y)}{2} \right) \right\},$$

where $x_j = \langle x, \varphi_j \rangle$ for all $j \geq 1$. This leads to the following estimation scheme for $f(x/y)$:

1. Obtain an estimate $\widehat{\mu}(\cdot/y)$ of $t \to \mu(t/y)$ for all $y \in \mathbb{R}$. This may be carried out trough any standard nonparametric regression from $\mathbb{R}$ to $\mathbb{R}$, based on the learning set $(y_i, x_i(t))_{i=1,\ldots,n}$; e.g., a smoothing kernel method.
2. Obtain estimates $(\widehat{\varphi}_j,\ \widehat{\lambda}_j)_j$ of the eigenfunctions and eigenvalues $(\varphi_j,\ \lambda_j)_j$ of the covariance r on the basis of the empirical covariance of the residuals $x_i - \widehat{\mu}(\cdot/y_i)$, $i = 1,...,n$. Only the first p eigenvalues and eigenfunctions are estimated, where $p = p(n)$ is a given integer, smaller than n.
3. Estimate $f(x/y)$ by

$$\widehat{f}(x/y) = \exp\left\{ \sum_{j=1}^{p} \frac{\widehat{\mu}_j(y)}{\widehat{\lambda}_j} \left(\widehat{x}_j - \frac{\widehat{\mu}_j(y)}{2} \right) \right\} \tag{3}$$

where $\widehat{\mu}_j(y) = \langle \widehat{\mu}(\cdot/y), \widehat{\varphi}_j \rangle$ and $\widehat{x}_j = \langle x, \widehat{\varphi}_j \rangle$.

Finally, substituting (3) into (2) leads to an estimate $\widehat{\gamma}(x)$ of $\gamma(x)$. Under some technical assumptions the consistency of the DBNR method can be proved: $\lim_{n\to\infty} \hat{\gamma}(x) =^{\mathbb{P}} \gamma(x)$.

3 A simulation study

The feasibility and the performance of the introduced nonparametric functional regression method are first explored through a simulation study. For comparison, results obtained by the functional Nadaraya-Watson kernel (NWK) estimator [10] are also shown.

3.1 Data generation

The data were simulated in the following way: values for the real random variable, Y, were drawn from a uniform distribution in the interval $[0, 10]$. Then, X was generated by 4 different models or settings:

M1 $X = Ye_1 + 2Ye_2 + 3Ye_5 + 4Ye_{10} + \epsilon$
M2 $X = (\exp(Y)/\exp(10))e_1 + (Y^2/100)e_2 + (Y^3/1000)e_5 + \log(Y+1)e_{10} + \epsilon$
M3 $X = \sin(Y)e_1 + \log(Y+1)e_5 + \epsilon$
M4 $X = \alpha \exp\left(\frac{Y}{10}e_1\right) + \epsilon$

where $(e_i)_{i\geq 1}$ is the trigonometric basis of $\mathcal{X} = \mathcal{L}^2([0,1])$ (i.e., $e_{2k-1} = \sqrt{2}\cos(2\pi kt)$, and $e_{2k} = \sqrt{2}\sin(2\pi kt)$), and ϵ a Gaussian process independent of Y with zero mean and covariance operator $\Gamma_e = \sum_{j\geq 1} \frac{1}{j} e_j \otimes e_j$. More precisely, ϵ was simulated by using a truncation of Γ_e, $\Gamma_e(s,t) \simeq \sum_{j=1}^{q} \frac{1}{j} e_j(t) e_j(s)$ with $q = 500$.

A sample of size $n_L = 300$ was simulated for training and a sample of size $n_T = 200$ for testing. Figure 1 gives examples of X obtained for model **M3** for three different values of y and of the underlying (non noisy) function, $F(y)(\cdot)$. In this example, the simulated data have a high level of noise so that the regression estimation is a rather hard statistical task.

3.2 Simulation results

To apply the DBNR method, the discretized functions X were approximated by a continuous function using a functional basis expansion. Specifically, the data were approximated using 128 B-spline basis functions of order 4, as it is shown in Figure 1. The conditional mean $\mu(\cdot/y)$ was estimated by a kernel smoothing in which the bandwidth parameter h was selected by 10-fold cross-validation minimizing the mean squared error (MSE) criterion. A similar procedure was used to select the parameter p (number of eigenvalues and eigenfunctions used in (3)).

Finally, DBNR performance was compared with those obtained by the functional NWK estimate with two kinds of metrics for the kernel: the usual $\mathcal{L}^2$-norm

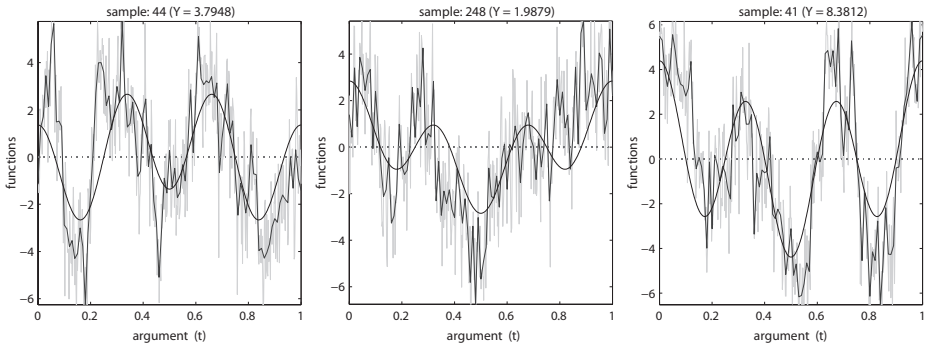

Fig. 1. True function, $F(y)(\cdot)$ (smooth continuous line), simulated data, X, (gray rough line) and approximation of X using B-splines (rough black line) in **M3** for three different values of y

and the PCA based semi-metric norm (see [10] for further details about these methods). The resulting root mean squared errors (RMSE) are presented in Table 1. The results show that DBNR is a good alternative to common NWK

Table 1. RMSE for all the methods and all generating models

Model	FGIR	NWK (PCA)	NWK ($\mathcal{L}^2$)
M1	0.08	0.10	0.09
M2	1.47	1.60	1.77
M3	1.79	1.79	2.00
M4	0.94	2.16	1.91

methods. Indeed, DBNR outperforms NWK methods in all the the cases considered in this simulation study that includes both linear (**M1**) and nonlinear (**M2** – **M4**) models.

Figures 2 and 3 show how the method performs for each step of the estimation scheme (described in Section 2.2) for the model **M3**. In particular, Figure 2 gives the result of the first step by displaying the true value and the estimate of $F(y)(\cdot)$ for various values of y (top) and the true value and the estimate of $F(\cdot)(t)$ for various values of t (bottom). The results are very satisfactory given the fact that the data have a high level of noise (which is stressed on in the bottom of the figure): a minor estimation problem appears at the boundaries of $F(\cdot)(t)$, which is a known drawback of the kernel smoothing method. Also, those estimates are smoother than the estimates of $F(y)(\cdot)$: this can be explained by the fact that the kernel estimator is used regarding y and not regarding t, but this aspect can be improved in the future.

Figure 3 shows the results of the steps 2-3 of the estimation scheme: the estimated eigendecomposition of r is compared to the true one and finally, the

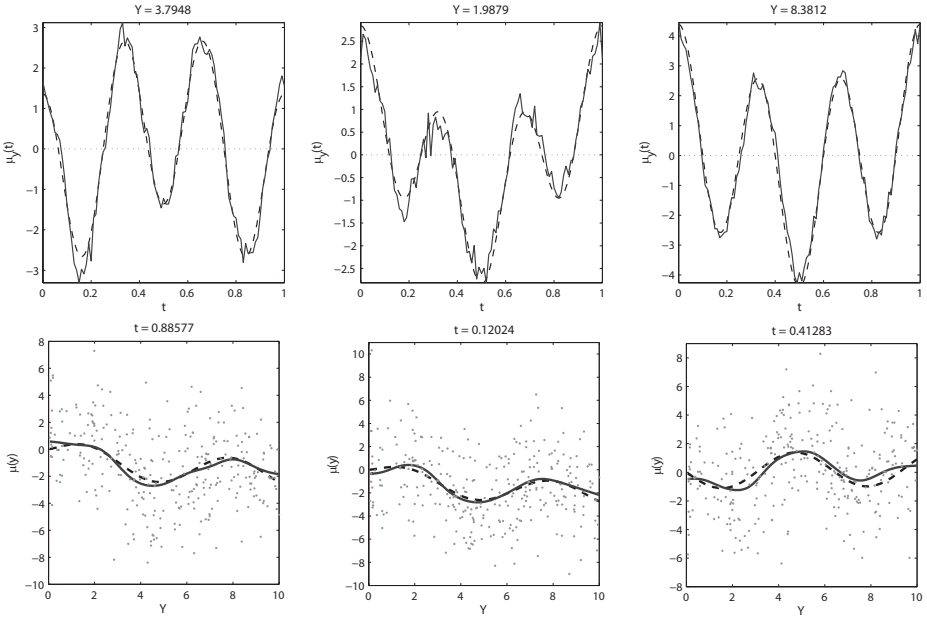

Fig. 2. True value (discontinuous lines) and estimate (continuous lines) of $F(y)(\cdot)$ for various values of y (top) and true value and estimate of $F(\cdot)(t)$ for various values of t (bottom) in model **M3**. The dots (bottom) are the simulated data, $X(t)$.

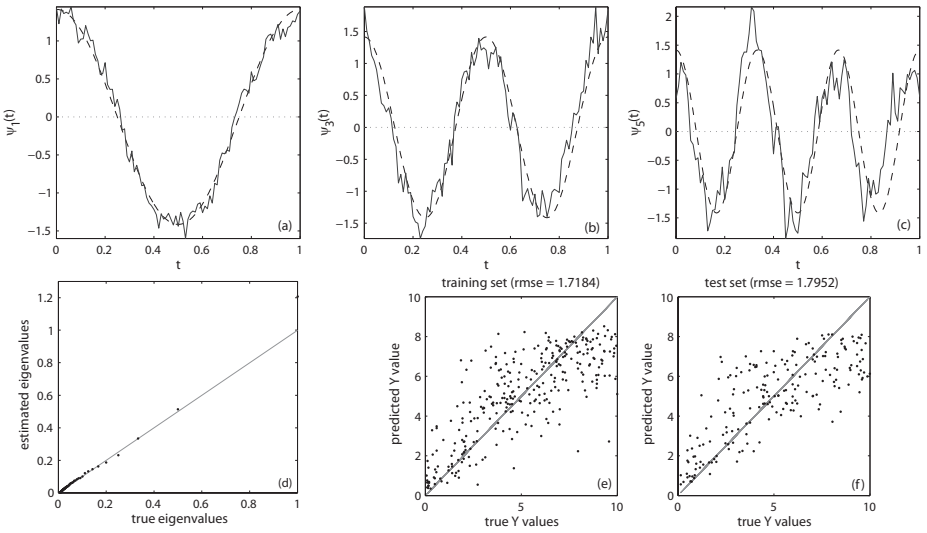

Fig. 3. Model **M3**: (a-c)True (dashed line) and estimated eigenfunctions (continuous line), (d) true and estimated eigenvalues and (d-e) predicted vs. true Y values for training and test sets.

predicted value for Y are compared to the true ones, both on training and test sets. The estimation of the eigendecomposition is, once again very satisfactory given the high level of noise, and the comparison between training and test sets show that the method does not overfit the data.

4 A study of Tecator dataset

DBNR was also tested on a benchmark data set for functional data: the Tecator dataset[1]. It consists in spectrometric data from the food industry. Each of the 215 observations is the near infrared absorbance spectrum of a meat sample recorded on a Tecator Infratec Food and Feed Analyzer. Each spectrum is sampled at 100 wavelengths uniformly spaced in the range 850–1050 nm. The composition of each meat sample is determined by analytic chemistry, so percentages of moisture, fat and protein are associated in this way to each spectrum. This problem is more challenging than the one presented in Section 3 where the data were generated to fulfill exactly the conditions of the DBNR model.

The whole data set was randomly split 100 times into training and test sets of almost the same size. The splits were randomly built such that also the training and test set were equally represented over the whole range of fat content.

Table 2 reports the mean of the MSE (and its standard deviation) over the 100 divisions both for DBNR and NWK methods.

Table 2. Prediction results on Tecator dataset

Model	DBNR	NWK (PCA)	NWK ($\mathcal{L}^2$)
MSE	1.91 (0.41)	9.1 (2.1)	8.9 (2.1)

Results obtained on Tecator by DBNR are the best in the sense of minimum MSE among all the methods. In [10] results based on the use of a semi-metric involving the second order derivatives (which is known to be useful for this data set) were also reported. A MSE of 3.5 was also obtained, which is still larger than the use of DBNR without derivative information.

5 Conclusions

A new functional nonparametric regression approach has been introduced motivated by the calibration problems in chemometrics. The new method, named functional density-based nonparametric regression (DBNR) was fully described under a Gaussian assumption for the distribution of X given Y but it could be extended to other kinds of distributions. The simulation study and the application of DBNR to a real data set have shown that DBNR performs well and outperforms functional NWK regression methods. Thus, DBNR can be considered

[1] Data are available on statlib at http://lib.stat.cmu.edu/datasets/tecator; see [18].

a promising alternative to existing functional regression methods, particularly appealing for calibration problems.

References

1. Ramsay, J., Silverman, B.: Functional Data Analysis. Second edn. Springer, New York (2005)
2. Ramsay, J., Dalzell, C.: Some tools for functional data analysis. Journal of the Royal Statistical Society, Series B **53** (1991) 539–572
3. Hastie, T., Mallows, C.: A discussion of a statistical view of some chemometrics regression tools by i. e. frank and j. h. friedman. Technometrics **35** (1993) 140–143
4. Marx, B.D., Eilers, P.H.: Generalized linear regression on sampled signals and curves: a p-spline approach. Technometrics **41** (1999) 1–13
5. Cardot, H., Ferraty, F., Sarda, P.: Functional linear model. Statistics and Probability Letter **45** (1999) 1122
6. Cardot, H., Ferraty, F., Sarda, P.: Spline estimators for the functional linear model. Statistica Sinica **13** (2003) 571591
7. Cardot, H., Crambes, C., Kneip, A., Sarda, P.: Smoothing spline estimators in functional linear regression with errors in variables. Comput. Statist. Data Anal. **51** (2007) 4832–4848
8. Crambes, C., Kneip, A., Sarda, P.: Smoothing splines estimators for functional linear regression. The Annals of Statistics (2008)
9. Preda, C., Saporta, G.: Pls regression on stochastic processes. Comput. Statist. Data Anal. **48** (2005) 149–158
10. Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)
11. Rossi, F., Conan-Guez, B.: Functional multi-layer perceptron: a nonlinear tool for functional data anlysis. Neural Networks **18**(1) (2005) 45–60
12. Preda, C.: Regression models for functional data by reproducing kernel hilbert space methods. J. Stat. Plan. Infer. **137** (2007) 829–840
13. Hernández, N., Biscay, R.J., Talavera, I.: Support vector regression methods for functional data. Lecture Notes in Computer Science **4756** (2008) 564–573
14. Li, K.: Sliced inverse regression for dimension reduction. J. Am. Stat. Assoc. **86** (1991) 316–327
15. Dauxois, J., Ferré, L., Yao, A.: Un modèle semi-paramétrique pour variable aléatoire hilbertienne. C.R. Acad. Sci. Paris **327(I)** (2001) 6947–952
16. Ferré, L., Yao, A.: Functional sliced inverse regression analysis. Statistics **37** (2003) 475488
17. Ferré, L., Villa, N.: Multi-layer perceptron with functional inputs : an inverse regression approach. Scandinavian Journal of Statistics **33** (2006) 807–823
18. Thodberg, H.: A review of bayesian neural network with an application to near infrared spectroscopy. IEEE Transaction on Neural Networks **7**(1) (1996) 56–72