Contribute 39
Model-free prediction with application to functional data analysis

Dimitris N. Politis

Abstract We show how the Model-Free Prediction Principle of Politis (2013) can be applied to nonparametric regression with a univariate response and regressors that are high-dimensional or even infinite-dimensional, i.e., functional. Without assuming an additive model with i.i.d. errors, the Model-Free Principle is capable of yielding both consistent predictors as well as prediction intervals.

39.1 Introduction

Consider regression data of the type $(Y_1, x_1), \ldots, (Y_n, x_n)$ where Y_k is the response associated with a regressor value x_k. Based on such data, statistical inference can be of two flavors: (a) *explaining/modeling the world*, and (b) *predicting a future state of the world*. In step (a), the issue is to discover and describe the relationship between the response variable Y to the regressor variable x. Step (b) amounts to predicting a yet unobserved response Y_{n+1} associated with a regressor value x_{n+1}. If the modeling step (a) has been accomplished, then the prediction problem can be solved by using the fitted model as if it were exact.

Since fitting a model gives the practitioner the ability to predict future responses one can ask if the converse is also true. The answer is yes: if one is able to predict the future response that is associated with any regressor value x, then an implied model-fitting is taking place as the curve explaining/predicting Y on the basis of x is being constructed.

But how can one predict without a model? The *Model-Free Prediction Principle* of Politis [2] substitutes the notion of *transformation* in place of a model, and places the emphasis on observable quantities, i.e., current and future data, as opposed to unobservable model parameters and estimates thereof. To briefly describe it, consider the vector of responses $\underline{Y}_m = (Y_1, \ldots, Y_m)'$. Thus, $\underline{Y}_n$ con-

Dimitris N. Politis
University of California–San Diego, USA, email: dpolitis@ucsd.edu

tains the already observed responses while $\underline{Y}_{n+1}$ contains $\underline{Y}_n$ plus the future (yet unobserved) response Y_{n+1} associated with regressor value x_{n+1}.

The Model-Free (MF) Prediction Principle amounts to using the structure of the problem—that also utilizes the regressors—in order to find an *invertible transformation* H_m that can map the vector $\underline{Y}_m$ to a vector $\underline{\epsilon}_m = (\epsilon_1, \ldots, \epsilon_m)'$ that has i.i.d. components *conditionally* on the regressor values $x_1, \ldots, x_m$; here m could be taken equal to either n or $n+1$ as needed. Note that the functional form of H_m is allowed to depend on the regressor values $x_1, \ldots, x_m$ although this is not explicitly denoted. Letting H_m^{-1} denote the inverse transformation, we then have $\underline{\epsilon}_m = H_m(\underline{Y}_m)$ and $\underline{Y}_m = H_m^{-1}(\underline{\epsilon}_m)$, i.e.,

$$\underline{Y}_m \stackrel{H_m}{\longmapsto} \underline{\epsilon}_m \text{ and } \underline{\epsilon}_m \stackrel{H_m^{-1}}{\longmapsto} \underline{Y}_m. \tag{39.1}$$

If the practitioner is successful in identifying the transformation H_m, then the prediction problem is reduced to the trivial one of predicting i.i.d. variables. To see why, note that eq. (39.1) with $m = n+1$ yields $\underline{Y}_{n+1} = H_{n+1}^{-1}(\underline{\epsilon}_{n+1}) = H_{n+1}^{-1}(\underline{\epsilon}_n, \epsilon_{n+1})$. But $\underline{\epsilon}_n$ can be treated as known given the data $\underline{Y}_n$; just use eq. (39.1) with $m = n$. Since the unobserved Y_{n+1} is just the $(n+1)^{th}$ coordinate of vector $\underline{Y}_{n+1}$, it follows that Y_{n+1} can also be expressed as a function of the unobserved ϵ_{n+1} (given the additional regressor value x_{n+1} of interest). Finally, note that predicting a function, say $g(\cdot)$, of an i.i.d. sequence $\epsilon_1, \ldots, \epsilon_n$ is straightforward since $g(\epsilon_1), \ldots, g(\epsilon_n)$ is simply another i.i.d. sequence.

Under regularity conditions, such a transformation H_m always exists although it is not unique. The challenge to the skills and expertise of the statistician is to be able to devise and estimate a workable such transformation for the problem at hand. In what follows, we show how this task can be accomplished in the nonparametric regression paradigm where the regressor x takes values in a high-dimensional or even a function space.

39.2 Nonparametric regression models

Throughout the paper, we consider regression data $(Y_1, x_1), \ldots, (Y_n, x_n)$ where Y_k is the *univariate* response associated with a regressor value x_k that takes values in a linear vector space $\mathbf{E}$ equipped with a semi-metric d. The space $\mathbf{E}$ can be high-dimensional or even infinite-dimensional, e.g., a function space; see Chapter 5 of Ferraty and Vieu [1] for details.

The regressors $x_1, \ldots, x_n$ are either assumed deterministic, or represent a realization of the random variables $X_1, \ldots, X_n$. In the latter case, it is often assumed that

$$(Y_j, X_j) \text{ for } j = 1, \ldots, n \text{ are i.i.d.} \tag{39.2}$$

The above is a vague structural assumption, and does not constitute a nonparametric model *per se*.

In the case of deterministic regressors, two popular additive models for nonparametric regression are given by

$$Y_j = \mu(x_j) + \varepsilon_j \text{ for } j = 1, \ldots, n \tag{39.3}$$

and

$$Y_j = \mu(x_j) + \sigma(x_j)\varepsilon_j \text{ for } j = 1, \ldots, n \tag{39.4}$$

where $\mu(x) = E(Y_j|X_j = x)$, $\sigma^2(x) = Var(Y_j|X_j = x)$, and the errors ε_j are i.i.d. $(0, \sigma^2)$; in the case of (39.4) it is assumed that $\sigma^2 = 1$ for identifiability.

The above two models have their analogs in the random design case, namely

$$Y_j = \mu(X_j) + \varepsilon_j \text{ for } j = 1, \ldots, n \tag{39.5}$$

and

$$Y_j = \mu(X_j) + \sigma(X_j)\varepsilon_j \text{ for } j = 1, \ldots, n \tag{39.6}$$

under the typical additional assumption that the i.i.d. errors $(\varepsilon_1, \ldots, \varepsilon_n)$ are independent of $(X_1, \ldots, X_n)$.

39.3 Model-Free regression with functional data

The term 'Model-free' refers to the absence of a model equation such as (39.5) or (39.6). The pairwise i.i.d. assumption (39.2) is only a vague structural assumption, and therefore qualifies to be called 'Model-free'. Throughout the rest of the paper, we will work with an even weaker version of (39.2) that is described in the next paragraph.

Model-free set-up. *The dataset is* $\{(Y_t, x_t),\ t = 1, \ldots, n\}$ *where the* **E***–valued regressors* $x_1, \ldots, x_n$ *are either deterministic, or represent a realization of the random variables* $X_1, \ldots, X_n$*. In the latter case, it will be assumed that* Y_j *is independent of* $\{X_k$ *for* $k \neq j\}$*, and inference will be conducted conditionally on event* $S_n = \{X_j = x_j$ *for* $j = 1, \ldots, n\}$*. Conditionally on* S_m*, for any* $m \geq 1$*, the responses* $Y_1, \ldots, Y_m$ *will be assumed independent although not identically distributed. Also assume that the conditional distribution* $P\{Y_j \leq y|X_j = x\}$ *does not depend on* j*.*

Remark 39.1. *In the case of random design, the above Model-free set-up implies* (39.2) *if one additionally assumes that* $X_1, \ldots, X_n$ *are i.i.d.*

For nonparametric estimation, some smoothness assumption is typically needed. We will work under the simple assumption that the common conditional distribution $D_x(y) = P\{Y_j \leq y|X_j = x\}$ is *continuous* in both x and y. Consequently, we can estimate $D_x(y)$ by the 'local' weighted average

$$\hat{D}_x(y) = \sum_{i=1}^{n} \mathbf{1}\{Y_i \leq y\}\tilde{K}\left(\frac{d(x, x_i)}{h}\right) \tag{39.7}$$

where $\tilde{K}\left(h^{-1}d(x, x_i)\right) = K\left(h^{-1}d(x, x_i)\right)/\sum_{k=1}^{n} K\left(h^{-1}d(x, x_k)\right)$, the kernel K is a bounded, symmetric probability density with compact support, and $h > 0$ is a bandwidth parameter.

For any fixed y, estimator $\hat{D}_x(y)$ is just a Nadaraya-Watson smoother of the variables $\mathbf{1}\{Y_i \leq y\}$ for $i = 1, \ldots, n$. As such, it is discontinuous as a function of y;

to come up with a continuous estimator, we can replace $\mathbf{1}\{Y_i \leq y\}$ by $\Lambda\left(\frac{Y_i - y}{b}\right)$ in eq. (39.7), leading to the estimator

$$\bar{D}_x(y) = \sum_{i=1}^{n} \Lambda\left(\frac{Y_i - y}{b}\right) \tilde{K}\left(\frac{d(x, x_i)}{h}\right) \tag{39.8}$$

where b is another bandwidth parameter, and $\Lambda(y) = \int_{-\infty}^{y} \lambda(s)ds$ with $\lambda(\cdot)$ being a symmetric density function that is continuous and strictly positive over its support. As a result, $\bar{D}_x(y)$ is continuous and strictly increasing in y.

Under model (39.2) and additional regularity conditions, e.g., that as $n \to \infty$, $\max(h, b) \to 0$ but not too fast, Theorem 6.4 of Ferraty and Vieu [1] shows

$$\bar{D}_x(y) \xrightarrow{a.s.} D_x(y) \text{ for any } y, \text{ and } \bar{D}_x^{-1}(\alpha) \xrightarrow{a.s.} D_x^{-1}(\alpha) \tag{39.9}$$

for any $\alpha \in [0, 1]$ as long as $D_x(y)$ is strictly increasing at $y = D_x^{-1}(\alpha)$. It is conjectured that a similar consistency result can be obtained in the case of deterministic regressors that follow a regular design.

39.4 Model-Free prediction with functional data

Conditionally on S_n, the Y_ts are non–i.i.d. but this is only because they do not have identical distributions. Since they are continuous random variables, the *probability integral trasform* is the key idea to transform them towards 'i.i.d.–ness'. To see why, note that if we let

$$\eta_i = D_{x_i}(Y_i) \quad \text{for } \; i = 1, \ldots, n$$

our goal would be exactly achieved since $\eta_1, \ldots, \eta_n$ are i.i.d. Uniform(0,1). Of course, $D_x(\cdot)$ is not known but we have the consistent estimator $\bar{D}_x(\cdot)$ as its proxy. Therefore, our proposed transformation amounts to defining

$$u_i = \bar{D}_{x_i}(Y_i) \quad \text{for } \; i = 1, \ldots, n. \tag{39.10}$$

Eq. (39.8) then implies that $u_1, \ldots, u_n$ are approximately i.i.d. Uniform(0,1).

We can now invoke the Model-Free Prediction Principle of Politis [2] in order to construct optimal predictors of $g(Y_{n+1})$ where Y_{n+1} is the out-of-sample response associated with regressor value x_{n+1}, and $g(\cdot)$ is any measurable function. The L_2–optimal predictor of $g(Y_{n+1})$ is the expected value of $g(Y_{n+1})$ given x_{n+1} that is estimated by

$$\Pi_2 = n^{-1} \sum_{i=1}^{n} g\left(\bar{D}_{x_{n+1}}^{-1}(u_i)\right). \tag{39.11}$$

Similarly, the L_1–optimal predictor of $g(Y_{n+1})$ suggested by the Model-Free Prediction Principle is $\Pi_1 =$ sample median of the set $\{g\left(\bar{D}_{x_{n+1}}^{-1}(u_i)\right),\ i = 1, \ldots, n\}$.

For simplicity, focus on the case where $g(y) = y$, and note that one can construct alternative estimators of the L_2 and L_1–optimal predictors of Y_{n+1}; these are respectively given by

$$\pi_2 = \sum_{i=1}^{n} Y_i \tilde{K}\left(h^{-1} d(x_{n+1}, x_i)\right) \text{ and } \pi_1 = \bar{D}^{-1}_{x_{n+1}}(1/2).$$

Eq. (39.9) shows that π_1 is a consistent estimator of the theoretical L_1–optimal predictor $D^{-1}_{x_{n+1}}(1/2)$. Under some additional regularity conditions, Ferraty and Vieu [1] also show that π_2 is consistent for $E(Y_{n+1}|X_{n+1} = x_{n+1})$ under model (39.2).

The new predictors Π_2 and Π_1 look quite different from the traditional predictors π_2 and π_1 but, surprisingly, they turn out to be asymptotically equivalent. For example, $\Pi_1 = \text{median}\{\bar{D}^{-1}_{x_{n+1}}(u_i)\} = \bar{D}^{-1}_{x_{n+1}}(\text{median}\{u_i\}) \simeq \bar{D}^{-1}_{x_{n+1}}(1/2) = \pi_1$ since the u_is are approximately Uniform (0,1), and $\bar{D}^{-1}_{x_{n+1}}(\cdot)$ is strictly increasing. Similarly, for any distribution F, we can define the quantile-inverse of F, i.e., $F^{-1}(u) = \inf\{y \text{ such that } F(y) \geq u\}$; from the identity $\int y F(dy) = \int_0^1 F^{-1}(u)du$ it then follows that

$$\pi_2 = \int y \, \hat{D}_{x_{n+1}}(dy) = \int_0^1 \hat{D}^{-1}_{x_{n+1}}(u)du \simeq \int_0^1 \bar{D}^{-1}_{x_{n+1}}(u)du \simeq \Pi_2.$$

Remark 39.2. *All the aforementioned predictors are based on either the estimator $\bar{D}_{x_{n+1}}(\cdot)$ or $\hat{D}_{x_{n+1}}(\cdot)$ whose finite-sample accuracy crucially depends on the number of data pairs (Y_j, X_j) with regressor value that lies in the neighborhood of the point of interest x_{n+1}. If few (or none) of the regressors are found close to x_{n+1}, then nonparametric prediction will be highly inaccurate (or even impossible).*

Remark 39.3. *The original assumption that $D_x(y)$ is continuous in y can be relaxed; see Politis [2] for a discussion on how to deal with discrete responses. In brief, when $D_x(y)$ is not assumed continuous in y the smooth estimator $\bar{D}_x(y)$ is not useful, and this precipitates two main changes to the methodology: (a) the u_i are not defined from eq.* (39.8) *any longer— rather we generate $u_1, \ldots, u_n$ as i.i.d. Uniform(0,1); and (b) we use $\hat{D}^{-1}_x(u)$ instead of $\bar{D}^{-1}_x(u)$ at all instances.*

39.5 Model-Free bootstrap and prediction intervals

As already mentioned, the Model-Free Prediction Principle suggests the predictors Π_2 and Π_1 which are asymptotically equivalent to the traditional predictors π_2 and π_1 respectively. Nevertheless, the main advantage of the Model-Free, transformation-based approach is that it allows us to go *beyond* point prediction and obtain valid predictive distributions and intervals for Y_{n+1}. To do this, however, some kind of resampling procedure is necessary in order to also capture the variance due to estimation error, e.g., the error in using Π_2 (or π_2) instead of the true $E(Y_{n+1}|X_{n+1} = x_{n+1})$, etc. For example, consider the prediction interval

$$[\hat{D}^{-1}_{x_{n+1}}(\alpha/2),\ \hat{D}^{-1}_{x_{n+1}}(1-\alpha/2)] \tag{39.12}$$

given in eq. (5.10) of Ferraty and Vieu [1]; this interval is indeed asymptotically valid as it will contain Y_{n+1} with probability tending to the nominal $(1-\alpha)100\%$. However, interval (39.12) will be characterized by *under-coverage* in finite samples since the non-trivial variability in the estimated quantiles $\hat{D}^{-1}_{x_{n+1}}(\alpha/2)$ and $\hat{D}^{-1}_{x_{n+1}}(1-\alpha/2)$ is ignored.

Now, having mapped the responses $Y_1, \ldots, Y_n$ onto the approximately i.i.d. variables $u_1, \ldots, u_n$, it is natural to perform an i.i.d. bootstrap on the latter, and then transform back to obtain bootstrap pseudo-responses. This is the idea for the *Model-Free bootstrap* described in Section 2.6 of Politis [2]; in particular, the bootstrap algorithms given in Sections 4.4 and 4.5 of Politis [2] apply *verbatim* to the current set-up of nonparametric regression with univariate response and functional regressors.

Note that the Model-Free bootstrap is performed treating the design points $x_1, \ldots, x_n$ as fixed; hence it is akin to the well-known residual bootstrap available when a model such as (39.3) holds true. One may consider instead resampling pairs which is associated with the pairwise i.i.d. assumption (39.2). However, by resampling the i.i.d. pairs (Y_j, X_j) we run a great risk of obtaining a bootstrap pseudo-sample $\{(Y_j^*, X_j^*)$ for $j = 1, \ldots, n\}$ for which few of the X_j^* are found in the neighborhood of the point of interest x_{n+1}, thus making nonparametric estimation impossible in the bootstap world; see Remark 39.**2**. By contrast, the Model-Free bootstrap does not have this potential disadvantage. In addition, it is valid under a slightly weaker set of assumptions than eq. (39.2); see Remark 39.**1**.

Finally, it is interesting that the Model-Free bootstrap can also be used to yield confidence bands for the conditional expectation and conditional variance functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ without assuming an additive model such as (39.5) or (39.6); the algorithms given in Politis [3] apply *verbatim* to the case of functional regressors.

Acknowledgement. The author is grateful to Anirban DasGupta, Stathis Paparoditis, and Philippe Vieu for helpful discussions.

Bibliography

[1] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*, Springer, New York.

[2] Politis, D.N. (2013). Model-free model-fitting and predictive distributions, (with Discussion), *Test*, vol. 22, no. 2, pp. 183-250.

[3] Politis, D.N. (2014). Bootstrap confidence intervals in nonparametric regression without an additive model, in *Proceedings of the First Conference of the International Society for NonParametric Statistics*, M.G. Akritas, S.N. Lahiri and D.N. Politis (Eds.), Springer, New York.