Journal of the Chinese Statistical Association
Vol. 54, (2016) 186–204

ON FINITE DIMENSIONAL APPROXIMATION IN NONPARAMETRIC REGRESSION

Wen Hsiang Wei†1
Department of Statistics, Tung Hai University, Taiwan

ABSTRACT

For the function from a real separable Banach space into a real separable Banach space, i.e., a possibly nonlinear operator, in nonparametric regression, theoretical results are established for the estimator based on finite dimensional approximation. A new concept "approximatability" is presented and the operators of interest are proved to be approximatable under different situations. The results concerning both consistency and weak convergence of the estimator are obtained. Statistical applications of these theoretical results are given.

Key words and phrases: Approximatability, Banach spaces, Consistency, Nonlinear operator, Nonparametric regression, Weak convergence.
JEL classification: C14, C61

†Correspondence to: Wen Hsiang Wei
E-mail:wenwei@thu.edu.tw

1. Introduction

Consider the following nonparametric regression model,

$$y = F(x) + \epsilon,$$

where F is usually a real-valued function on R^q, $x \in R^q$, and both y and ϵ are real-valued random variables.

The above nonparametric regression model can be generalized by relaxing the assumptions imposed on the domain and range of the function F. In this article, let F defined on some subset of X be a Y-valued function and fall in a real separable Banach space V with a known Schauder basis (Enflo, 1973; Kreyszig, 1978, p. 69; Morrison, 2001, Proposition 5.3) and both y and ϵ are Y-valued random variables, where both X and Y are real separable Banach spaces. Relatively few theoretical results have been done for the resulting estimators based on finite dimensional approximations in such general setting. The goal of this article is to develop the theoretical results for the "generalized" nonparametric regression models. The unknown operator $F = \sum_{j=1}^{\infty} \beta_j \varphi_j$ is of interest, where $\{\varphi_j\}$ is the Schauder basis and β_j are the coefficients. To estimate F, the coefficients β_j need to be estimated based on the data available first. In practice, the finite dimensional approximations of F, i.e., $\sum_{j=1}^{m} \beta_j \varphi_j$, can be used as the objective operator to be estimated. Thus, the methods, usually involving some numerical algorithms, can be employed to estimate the finite number of coefficients. As the resulting estimator $\sum_{j=1}^{m} \hat{\beta}_{nj} \varphi_j$ is an accurate estimator of $\sum_{j=1}^{m} \beta_j \varphi_j$, for example, the consistent estimator, it can be also an accurate estimator of F for large m due to $\sum_{j=m+1}^{\infty} \beta_j \varphi_j \underset{m \to \infty}{\longrightarrow} 0$, where $\hat{\beta}_{nj}$ are real-valued random variables and n is a positive integer, usually the number of data available. In next section, Theorem 2.1 and its associated corollaries indicate that the estimated operator (the minimizer) of a convex objective functional based on the finite dimensional approximations of V might converge to the one based on the original infinite dimensional space in different situations. In Section 3, the consistency of the estimator $\sum_{j=1}^{m} \hat{\beta}_{nj} \varphi_j$ is proved. Furthermore, some sufficient conditions for the convergence of the sequence of estimators in distribution to a centered Radon Gaussian variable (Ledoux & Talagrand, 1991, Chapter 3) are established in this section. Finally, the theoretical results given in Section 2 and Section 3 can be employed in a variety of statistical models, including the nonparametric regression models with or without measurement errors and the models for fitting the functional data with or without the use of a differential equation. These statistical applications are presented in Section 4. Hereafter, all the normed spaces or the inner product spaces of interest are over the real field and the notation $|| \cdot ||_W$ is denoted as the norm of the normed space W. As W is a Hilbert space, the norm induced by the inner product is $|| \cdot ||_W = (< \cdot, \cdot >_W)^{1/2}$.

2. Finite Dimensional Approximation

The concept of finite dimensional approximation is described first.

Definition 2.1. *Let S be a non-empty subset of a normed space U. u in S is approximatable by a sequence $\{u_m : u_m \in U_m \cap S\}$ if the sequence satisfies*

$$u_m \underset{m\to\infty}{\longrightarrow} u,$$

where U_m are finite dimensional subspaces of U. An operator $T : S \to W$ taking values on a normed space W is approximatable at u by $\{u_m\}$ if

$$T(u_m) \underset{m\to\infty}{\longrightarrow} T(u).$$

As $S = V$, any element u in S and any continuous operator T on S can be approximatable by the sequence of which elements are linear combinations of the Schauder basis vectors. However, the approximatability result might not be true if the set S is not the space V or T is not continuous on S. It turns out that the convexity of the set S and the functional T, i.e., $W = R$, plays a crucial role for the approximatability of the statistical estimator .

A commonly used method to estimate true F is to find the minimizer(s) of some objective functional, for example, the minimizer of the residual sum of squares. Let S be a non-empty closed convex subset of V and $T : S \to R$ be convex, lower semi-continuous, and proper, where V is assumed to be reflexive in this section. As S is unbounded, T is assumed to be coercive. The existence of the minimizer(s) of a certain objective functional on V is well established (Deimling, 1985, Theorem 25.1; Ekeland & Témam, 1999, Proposition 1.2, p. 35). By Proposition 1.2 of Ekeland & Téman (1999), the minimizer $\hat{F} \in S$ of the objective functional T exists as indicated by the following lemma.

Lemma 2.1. $\hat{F} = \underset{F\in S}{\arg\min}\, T(F)$ *exists. Furthermore, $\hat{F}$ is unique if T is strictly convex on S.*

Based on the above lemma, the finite dimensional approximation of $\hat{F}$ exists as indicated by the following theorem.

Theorem 2.1. *Let $\hat{F} \in S^\circ$, where S° is the interior of S. There exist subspaces V_m of V spanned by the finite number of elements of the Schauder basis $\{\varphi_j\}$ and a sequence of minimizers $\{\hat{F}_m = \underset{F\in V_m\cap S}{\arg\min}\, T(F)\}$ such that T is approximatable at $\hat{F}$ by $\{\hat{F}_m\}$. As T is strictly convex, the unique minimizer $\hat{F}$ is approximatable by $\{\hat{F}_m\}$.*

Proof. Because $\hat{F} \in S^{\circ}$, for every $\epsilon_1 = 1/m > 0$, there exists an N_1 depending on m such that $||\tilde{F}_m - \hat{F}||_V < \epsilon_1$ as $n_m > N_1$, where $\tilde{F}_m = \sum_{j=1}^{n_m} b_j \varphi_j \in S$. Note that $\{n_m : m = 1, 2, \ldots\}$ can be chosen to be a nondecreasing sequence. That is, $\tilde{F}_m \underset{m\to\infty}{\longrightarrow} \hat{F}$. Moreover, because T is continuous at $\hat{F}$ by Corollary 2.5 (p. 13) of Ekeland & Témam (1999), for every $\epsilon_2 > 0$, there exists an N_2 such that $|T(\tilde{F}_m) - T(\hat{F})| < \epsilon_2$ for $m > N_2$, i.e., $T(\tilde{F}_m) \underset{m\to\infty}{\longrightarrow} T(\hat{F})$. Let $V_m = span\{\varphi_j, j = 1, \ldots, n_m\}$. Since $V_m \cap S$ is a nonempty closed convex subset of V, $\hat{F}_m = \underset{F\in V_m\cap S}{\arg\min}\, T(F)$ exists by Proposition 1.2 of Ekeland & Téman (1999). If $\hat{F}_m = \tilde{F}_m$, then $T(\hat{F}_m) = T(\tilde{F}_m) \underset{m\to\infty}{\longrightarrow} T(\hat{F})$. On the other hand, if $\hat{F}_m \neq \tilde{F}_m$, let $T(\hat{F}_m) - T(\tilde{F}_m) = \Delta_m < 0$. Then,

$$0 \leq T(\hat{F}_m) - T(\hat{F}) = T(\hat{F}_m) - T(\tilde{F}_m) + T(\tilde{F}_m) - T(\hat{F}) \leq \Delta_m + \epsilon_2.$$

This implies $|\Delta_m| \leq \epsilon_2$ and hence $|T(\hat{F}_m) - T(\tilde{F}_m)| \leq \epsilon_2$. $T(\hat{F}_m) \underset{m\to\infty}{\longrightarrow} T(\hat{F})$ then. As T is strictly convex, $\hat{F}$ and $\hat{F}_m$ are both unique by the above proposition and hence $\hat{F}_m \underset{m\to\infty}{\longrightarrow} \hat{F}$.

$\diamondsuit$

The approximatability holds under different situations, including different choices of V and the objective functional T, as indicated by the following corollaries. The first corollary indicates that the above theorem holds for the objective functional T depending on the difference of F and some usually pre-specified operator F_0.

Corollary 2.1. *Let V be a separable Hilbert space and $\phi : R^+ \to R$ be a nondecreasing, convex, lower semi-continuous, coercive, and proper function, where $R^+ = \{t : t \geq 0, t \in R\}$.*
(a) If $T(F) = \phi(\|F - F_0\|_V)$, then $\hat{F}$ exists and T is approximatable at $\hat{F}$ by $\{\hat{F}_m\}$, where $F \in S$ and $F_0 \in V$.
(b) If ϕ is strictly increasing and $T(F) = \phi(||F - F_0||_V^p)$, $1 < p < \infty$, then $\hat{F}$ is unique and approximatable by $\{\hat{F}_m\}$.

Proof. Since V, the separable Hilbert space, is also a reflexive Banach space with a Schauder basis, it suffices to prove that T is convex or strictly convex, lower semi-continuous, coercive, and proper. Then, the conditions in Theorem 2.1 hold.

Because ϕ is coercive and proper, T is coercive and proper thus. Since the normed function is a convex function, then for $F_1, F_2 \in S$ and $0 \leq \alpha \leq 1$,

$$\begin{aligned}
& T\left[\alpha F_1 + (1-\alpha)F_2\right] \\
\leq\ & \phi\left[\alpha\|F_1 - F_0\|_V + (1-\alpha)\|F_2 - F_0\|_V\right] \\
\leq\ & \alpha\phi(\|F_1 - F_0\|_V) + (1-\alpha)\phi(\|F_2 - F_0\|_V) \\
=\ & \alpha T(F_1) + (1-\alpha)T(F_2).
\end{aligned}$$

Thus, T is convex. To prove lower semicontinuity of T, there exists a δ such that for every $\epsilon > 0$ and $z \in R^+$, $\phi(z) < \phi(z^*) + \epsilon$ as $|z^* - z| < \delta$. Thus, for any $F \in S$ and $||F^* - F||_V < \delta$, then

$$\begin{aligned} & T(F) \\ = \; & \phi(\|F - F_0\|_V) \\ < \; & \phi(\|F^* - F_0\|_V) + \epsilon \\ = \; & T(F^*) + \epsilon, \end{aligned}$$

owing to

$$|\|F^* - F_0\|_V - \|F - F_0\|_V| \leq \|F^* - F\|_V.$$

Therefore, T is lower semicontinuous on S.

As ϕ is strictly increasing, by strict convexity of $||\cdot||_V^p$ on S (Bauschke & Combettes, 2011, Example 8.21), $1 < p < \infty$,

$$\begin{aligned} & T\left[\alpha F_1 + (1-\alpha)F_2\right] \\ < \; & \phi\left[\alpha\|F_1 - F_0\|_V^p + (1-\alpha)\|F_2 - F_0\|_V^p\right] \\ \leq \; & \alpha T(F_1) + (1-\alpha)T(F_2), \end{aligned}$$

for $F_1, F_2 \in S$, $F_1 \neq F_2$, and $0 < \alpha < 1$, i.e., $T(F) = \phi(||F - F_0||_V^p)$ being strictly convex.

$\diamondsuit$

A Hilbert-Schmidt operator $F : X \to Y$ (Da Prato & Zabczyk, 1992, Appendix C) is a bounded linear operator with the norm

$$\|F\|_{HS} = \left(\sum_{i=1}^{\infty} \|F(e_i)\|_Y^2\right)^{1/2},$$

where X and Y are both separable Hilbert spaces, $\{e_i\}$ is the orthonormal basis of X, and $\|\cdot\|_Y$ is the norm induced by the inner product $< \cdot, \cdot >_Y$. The space consisting of Hilbert-Schmidt operators is a separable Hilbert space with the inner product $< F_1, F_2 >_{HS} = \sum_{i=1}^{\infty} < F_1(e_i), F_2(e_i) >_Y$ for F_1, F_2 in the space. The above corollary can be applied to the Hilbert-Schmidt operator involving random vectors.

Corollary 2.2. *Let $\phi : R^+ \to R$ be a nondecreasing, convex, lower semi-continuous, coercive, and proper function and V be the inner product space of Hilbert-Schmidt operators from X to Y with the norm $||\cdot||_{HS}$, where X is the separable Hilbert space* *and Y is the spaces of k-dimensional real random vectors of which elements are square-integrable, i.e., having the second moments, and $||y||_Y = [\sum_{i=1}^k E(y_i^2)]^{1/2}$ for $y = (y_1, \ldots, y_k)^t \in Y$. Let*

$$S = \{F : E\left[F(x)\right] = E\left[F_0(x)\right], F, F_0 \in V, x \in X\}.$$

(a) If $T(F) = \phi(\|F - F_0\|_{HS})$, then $\hat{F}$ exists and T is approximatable at $\hat{F}$ by $\{\hat{F}_m\}$.
(b) If ϕ is strictly increasing and $T(F) = \phi(||F - F_0||_{HS}^p)$, $1 < p < \infty$, $\hat{F}$ is unique and approximatable by $\{\hat{F}_m\}$.

Proof. S is convex because for $F_1, F_2 \in S$, $x \in X$, and $0 \le \alpha \le 1$,

$$E\left[\alpha F_1(x) + (1-\alpha)F_2(x)\right] = E\left[F_0(x)\right],$$

and hence $\alpha F_1 + (1-\alpha)F_2 \in S$. Next is to prove that S is closed. Let $F_n \underset{n\to\infty}{\longrightarrow} F$, i.e., $||F_n - F||_{HS} \underset{n\to\infty}{\longrightarrow} 0$, where $F_n \in S$ and $F \in V$. Because $||F_n - F|| \le ||F_n - F||_{HS}$ by Corollary 16.9 of Meise & Vogt (1997), $||F_n(x) - F(x)||_Y \underset{n\to\infty}{\longrightarrow} 0$, where $||\cdot||$ is the usual operator norm for the bounded linear operator. This gives that $E[F_n(x)] = E[F_0(x)] \underset{n\to\infty}{\longrightarrow} E[F(x)]$ and hence $E[F(x)] = E[F_0(x)]$. Hence, $F \in S$ and S is closed. The results follow by Corollary 2.1.

◇

Remark 2.1. *In linear regression, it is well known that the least squares estimate is also a BLUE (best linear unbiased estimate, Seber, 1977, Theorem 3.2). In fact, the existence of the BLUE is a special case of the above corollary. Let $\chi = (\chi_1, \ldots, \chi_p)^t$ be a p-dimensional random vector with $E(\chi) = \theta$ and the variance-covariance matrix equal to $\sigma^2 I$, where $\theta \in R^p$ and I is the identity matrix. Let the one-dimensional Hilbert space $X = \{a\chi : a \in R\}$ with the inner product $< a_1\chi, a_2\chi >_X = a_1 a_2 \sum_{i=1}^p E(\chi_i^2)$ and the orthonormal basis $e = \chi/||\chi||_X$, where $a_1, a_2 \in R$ and $||\cdot||_X$ is the norm induced by the inner product. Consider the bounded linear operators $F : X \to Y$ defined by $F(\chi) = l^t\chi$ and $F_0 : X \to Y$ defined by $F_0(\chi) = c^t E(\chi) = c^t\theta$, $l, c \in R^p$, where Y is the (p+1)-dimensional Hilbert space of real-valued random variables spanned by $\{\chi_i\} \cup \{1\}$ with the inner product $< y_1, y_2 >_Y = E(y_1 y_2)$ for $y_1, y_2 \in Y$ and the induced norm $||\cdot||_Y$. Note that both F and F_0 are Hilbert-Schmidt operators because $||F(e)||_Y < \infty$ and $||F_0(e)||_Y < \infty$. Further, the space V is a (p+1)-dimensional Hilbert space and hence a separable and reflexive Banach space. For any given c, the set S given in the above corollary consists of the "unbiased" operators with expected values equal to $c^t\theta$ at χ, i.e., $S = \{F : E[F(\chi)] = c^t\theta = E[F_0(\chi)]\}$. $T(F) = \phi(||F - F_0||_{HS}^2) = Var(l^t\chi)$ is strictly convex and hence there exists a unique minimizer, i.e., the BLUE with the minimum variance, where $\phi : R^+ \to R$ is given by $\phi(s) = ||\chi||_X^2 s, s \in R^+$.*

Reproducing kernel Hilbert space (Aronszajn, 1950) has been extensively used in nonparametric regression (Berlinet & Thomas-Agnan, 2004, Chapter 3). As the space V is a separable reproducing kernel Hilbert space, the following corollary can be applied to some optimization problems in nonparametric regression.

Corollary 2.3. *Let S be bounded and V be a separable reproducing kernel Hilbert space of real-valued functions defined on the subset of the separable Banach space X with a reproducing kernel $R(\cdot,\cdot)$ and an inner product $<\cdot,\cdot>_V$. Let $\phi_i : R \to R$ be convex, lower semi-continuous, and proper. Let $T(F) = \sum_{i=1}^{n} \phi_i\left[F(x_i)\right]$, where $R_{x_i} = R(\cdot, x_i)$ are not all equal to the zero element in V.*
(a) $\hat{F}$ exists and T is approximatable at $\hat{F}$ by $\{\hat{F}_m\}$.
(b) If S is the subset of S_R and ϕ_i are strictly convex, then $\hat{F}$ is unique and approximatable by $\{\hat{F}_m\}$, where S_R is the space spanned by $\{R_{x_i}\}$.

Proof. It suffices to prove that T is convex or strictly convex, lower semi-continuous, and proper. Then, the results follow by Theorem 2.1.

T is proper because ϕ_i is proper. Next is to prove convexity and lower semi-continuity of T. Because ϕ_i is lower semi-continuous, there exists a δ depending on F such that $\phi_i[F(x_i)] < \phi_i(s) + \epsilon/n$ for $|s - F(x_i)| < \delta$, every $\epsilon > 0$, and any $F \in S$. Thus, for $||F^* - F||_V < \delta/M$ and $M = \max_{1\le i\le n} ||R_{x_i}||_V$,

$$T(F) < T(F^*) + \epsilon,$$

because

$$\begin{aligned} & |F^*(x_i) - F(x_i)| \\ = & \ |< F^* - F, R_{x_i} >_V| \\ \le & \ \|F^* - F\|_V \|R_{x_i}\|_V \\ < & \ \delta, \end{aligned}$$

and hence $\phi_i[F(x_i)] < \phi_i[F^*(x_i)] + \epsilon/n$. Therefore, T is lower semi-continuous on S. Finally, because ϕ_i is convex and hence for $F_1, F_2 \in S$ and $0 \le \alpha \le 1$,

$$\begin{aligned} & T\left[\alpha F_1 + (1-\alpha)F_2\right] \\ = & \sum_{i=1}^{n} \phi_i\left[< \alpha F_1 + (1-\alpha)F_2, R_{x_i} >_V\right] \\ \le & \sum_{i=1}^{n} \left[\alpha\phi_i\left(< F_1, R_{x_i} >_V\right) + (1-\alpha)\phi_i\left(< F_2, R_{x_i} >_V\right)\right] \\ = & \ \alpha T(F_1) + (1-\alpha)T(F_2), \end{aligned}$$

T is convex.

Strict convexity of T given the conditions in (b) can be proved as follows. For $F_1, F_2 \in S$, $F_1 \neq F_2$, and $0 < \alpha < 1$, $T[\alpha F_1 + (1-\alpha)F_2] < \alpha T(F_1) + (1-\alpha)T(F_2)$ because $< F_1, R_{x_j} >_V \neq < F_2, R_{x_j} >_V$ for some j and then by strict convexity of ϕ_j,

$$\begin{aligned}&\phi_j\left[< \alpha F_1 + (1-\alpha)F_2, R_{x_j} >_V\right] \\ < \ & \alpha\phi_j\left(< F_1, R_{x_j} >_V\right) + (1-\alpha)\phi_j\left(< F_2, R_{x_j} >_V\right).\end{aligned}$$

Therefore, T is strictly convex on S.

$\diamondsuit$

Remark 2.2. *If $T(F) = \sum_{i=1}^{n} \phi_i\left(< F, v_i >_V\right)$, the same conditions imposed on S and ϕ_i in Corollary 2.3 hold, S_R is equal to the space spanned by $\{v_i\}$, and V, not necessarily being a reproducing kernel Hilbert space, is a separable Hilbert space, the results in the corollary still hold, where $v_i \in V$. Thus, even as F is a measurable function with respect to some measure, for example, F being a square-integrable function with respect to the Lebesgue measure, the results concerning the existence of $\hat{F}$ and the approximatability of $T(\hat{F})$ or $\hat{F}$ may still hold.*

3. Consistency and Weak Convergence of Nonlinear Estimators

Let $\hat{\beta}_n = (\hat{\beta}_{n1}, \ldots, \hat{\beta}_{nm})^t$ and $\beta = (\beta_1, \ldots, \beta_m)^t$, where $\hat{\beta}_{nj}$ are real-valued random variables. Let $\hat{\mathcal{G}}_n = \sum_{j=1}^{m} \hat{\beta}_{nj}\psi_j = \hat{\beta}_n^t\psi$ be the estimator of the operator $\mathcal{G} = \sum_{j=1}^{m} \beta_j\psi_j = \beta^t\psi$, where $\psi = [\psi_1, \ldots, \psi_m]^t$ and ψ_j are elements in some normed space W. The consistency and the asymptotic normality of the estimated coefficients $\hat{\beta}_n$ are key conditions for the convergence in probability and in distribution of the estimator. Denote the notations $\stackrel{p}{\longrightarrow}$ and $\stackrel{d}{\longrightarrow}$ as the convergence in probability and in distribution, respectively. If the consistency of the estimated coefficients $\hat{\beta}_n$ holds, the estimator based on the finite dimensional approximation is a consistent estimator of its counterpart, as indicated by the following theorem and corollary.

Theorem 3.1. If $\hat{\beta}_n \underset{n\to\infty}{\stackrel{p}{\longrightarrow}} \beta$, then $\hat{\mathcal{G}}_n \underset{n\to\infty}{\stackrel{p}{\longrightarrow}} \mathcal{G}$.

Proof. The measurability of $\hat{\mathcal{G}}_n$ is proved first. Define the function $h : R^m \to W$ by $h(x) = \sum_{j=1}^{m} x_j\psi_j$, where $x = (x_1, \ldots, x_m)^t$. h is continuous because

$$\|h(x_n) - h(x)\|_W \le m\left(\max_{1\le j\le m} \|\psi_j\|_W\right)\|x_n - x\|_{R^m} \underset{n\to\infty}{\longrightarrow} 0$$

as $\|x_n - x\|_{R^m} \underset{n\to\infty}{\longrightarrow} 0$. Therefore, $h(\hat{\beta}_n) = \hat{\mathcal{G}}_n$ is Borel measurable and thus a W-valued random variable. The result follows since

$$\left\|\hat{\mathcal{G}}_n(w) - \mathcal{G}(w)\right\|_W \le m\left(\max_{1\le j\le m} \|\psi_j\|_W\right)\left\|\hat{\beta}_n(w) - \beta\right\|_{R^m},$$

where w is any sample point.

$\diamondsuit$

As $W = V$, a direct result based on the above theorem is the following corollary. Let $\varphi = [\varphi_1, \ldots, \varphi_m]^t$.

Corollary 3.1. If $\hat{\beta}_n \xrightarrow[n\to\infty]{p} \beta$, then $\hat{\beta}_n^t\varphi \xrightarrow[n\to\infty]{p} \beta^t\varphi$.

The sufficient conditions for weak convergence of probability measures on the space of continuous functions defined on the unit interval $[0, 1]$ endowed with the uniform topology have been established (Billingsley, 1999, Theorem 7.1 and Theorem 7.5). Basically, the tightness of the sequence of probability measures and weak convergence of the finite dimensional distributions are two main conditions. The results have been generalized to the space of continuous operators endowed with the uniform topology, denoted by $C(K, Y)$, from a compact subset K of a separable Banach space X to a separable Banach space Y in Wei (2016). The result is given below. Let

$$w(F, \Delta) = sup_{\|x_1 - x_2\|_X \le \Delta} \|F(x_1) - F(x_2)\|_Y,$$

for $F \in C(K, Y)$.

Theorem 3.2. *Let $\{\mathcal{F}_n\}$ and $\mathcal{F}$ be $C(K, Y)$-valued random variables. If*

$$\lim_{\Delta\to 0} \limsup_{n\to\infty} P\left[w(\mathcal{F}_n, \Delta) \ge \epsilon\right] = 0,$$

and the sequence $\{[\mathcal{F}_n(x_1), \ldots, \mathcal{F}_n(x_k)]\}$ converges in distribution to $[\mathcal{F}(x_1), \ldots, \mathcal{F}(x_k)]$ for all $x_1, \ldots, x_k$ in K, i.e., the finite dimensional convergence of $\{\mathcal{F}_n\}$ in distribution to $\mathcal{F}$ in K, then $\{\mathcal{F}_n\}$ converges in distribution to $\mathcal{F}$.

The above theorem can be used for proving weak convergence of the operator-valued estimators in nonparametric regression, as indicated by the following theorem. Let $\hat{\beta}_{nj}$ be a $C(K^N, R)$-valued random variable, i.e., the assumption imposed on $\hat{\beta}_{nj}$ being relaxed. $\{\hat{\beta}_n(\tilde{x})\}$ is asymptotically normal on K^N if there exists a sequence of $m \times m$ matrices $\{c_n\}$ with $C(K^N, R)$-valued elements satisfying $\{\hat{\beta}_n^*(\tilde{x}) = [\hat{\beta}_{n1}^*(\tilde{x}), \ldots, \hat{\beta}_{nm}^*(\tilde{x})]^t = c_n(\tilde{x})(\hat{\beta}_n(\tilde{x}) - \beta)\}$ converges in distribution to a multivariate normal variable with zero mean vector and identity variance-covariance matrix for every $\tilde{x} \in K^N$. For a B-valued Radon Gaussian variable g, let

$$\Sigma(g) = \sup_{||T||_{B^*} \le 1, T \in B^*} \{E\{[T(g)]^2\}\}^{1/2},$$

where B^* is the topological dual space of the Banach space B. Assume that $0 \in K^{N+1}$, the range of the norm function on K^{N+1} is $[0, 1]$, and $0 \le ||x||_{X^{N+1}} \le 1$ for $x \in K^{N+1}$.

Theorem 3.3. *Let φ_j defined on K, $j = 1, \dots, m$, satisfy the Lipschitz condition, i.e., there exists a constant L_1 such that $||\varphi_j(\breve{x}) - \varphi_j(\breve{x}^*)||_Y \le L_1||\breve{x} - \breve{x}^*||_X$ for $\breve{x}, \breve{x}^* \in K$. Assume that there exist an n_0^*, positive constants L_2, δ such that*

$$P\left(\left|\hat{\beta}^*_{nj}(\tilde{x}) - \hat{\beta}^*_{nj}(\tilde{x}^*)\right| \le L_2\|\tilde{x} - \tilde{x}^*\|_{X^N}, \tilde{x}, \tilde{x}^* \in K^N\right) = 1, j = 1, \dots, m,$$

for $n \ge n_0^*$, and the number of points in a Δ-net for K^{N+1}, v, satisfies

$$v \le h(\Delta^{-1})\Delta^{-1},$$

where $\Delta < \delta$, h is an increasing function, and $h(n) = o[\exp(an^2)], \forall a > 0$. If $\{\hat{\beta}_n(\tilde{x})\}$ is asymptotically normal on K^N, then

$$\mathcal{F}_n = [c_n(\hat{\beta}_n - \beta)]^t\varphi \xrightarrow[n\to\infty]{d} \mathcal{F},$$

where $\mathcal{F} = \sum_{j=1}^m z_j\varphi_j$ is a centered Radon Gaussian variable with

$$\Sigma\left(\mathcal{F}\right) = sup_{\|T\|_{C^*(K,Y)}\le 1, T\in C^*(K,Y)}\left\{\sum_{j=1}^m \left[T\left(\varphi_j\right)\right]^2\right\}^{1/2},$$

and where $z_1, \dots, z_m$ are independent standard normal random variables.

Proof. First, $\{[\mathcal{F}_n(x_1), \dots, \mathcal{F}_n(x_k)]\}$ converges in distribution to

$$[\sum_{j=1}^m z_j\varphi_j(\breve{x}_1), \dots, \sum_{j=1}^m z_j\varphi_j(\breve{x}_k)]$$

by the asymptotical normality of $\{\hat{\beta}_n(\tilde{x})\}$ and the mapping theorem (Billingsley, 1999, Theorem 2.7), where $x_i = (\tilde{x}_i, \breve{x}_i)$, $\sum_{j=1}^m z_j\varphi_j(\breve{x}_i)$ is a Y-valued centered Radon Gaussian variable with

$$\Sigma\left[\sum_{j=1}^m z_j\varphi_j(\breve{x}_i)\right] = \sup_{\|T\|_{Y^*}\le 1, T\in Y^*}\left\{\sum_{j=1}^m \{T[\varphi_j(\breve{x}_i)]\}^2\right\}^{1/2},$$

for any $x_1, \dots, x_k$ in K^{N+1}. Thus, the finite dimensional convergence of $\{\mathcal{F}_n\}$ in distribution to $\mathcal{F}$ in K^{N+1} given in Theorem 3.2 holds.

Secondly, to prove that the equation in Theorem 3.2 holds, the following inequality is proved first,

$$P\left[w(\mathcal{F}_n, \Delta) \ge \epsilon_1\right] \le \sum_{i=1}^v P\left[\sup_{s\in\bar{B}(x_i, 2\Delta)} \|\mathcal{F}_n(s) - \mathcal{F}_n(x_i)\|_Y \ge \frac{\epsilon_1}{2}\right],$$

for given $\epsilon_1 > 0$, where $0 \leq ||x_1||_{X^{N+1}} \leq \cdots \leq ||x_v||_{X^{N+1}} \leq 1$, $\{x_i : i = 1, \ldots, v\}$ is a finite Δ-net for K^{N+1}, and $\bar{B}(x_i, 2\Delta)$ is the closed ball with the center x_i and the radius 2Δ. As $||s - x||_{X^{N+1}} \leq \Delta$, s and x fall in $\bar{B}(x_i, 2\Delta)$ because there exists a x_i such that $||x - x_i||_{X^{N+1}} < \Delta$ and hence

$$\|s - x_i\|_{X^{N+1}} \leq \|s - x\|_{X^{N+1}} + \|x - x_i\|_{X^{N+1}} < 2\Delta.$$

Then,

$$\|F(s) - F(x)\|_Y \leq \|F(s) - F(x_i)\|_Y + \|F(x_i) - F(x)\|_Y$$

and hence

$$w(F, \Delta) \leq 2 \max_{1 \leq i \leq v} \sup_{s \in \bar{B}(x_i, 2\Delta)} \|F(s) - F(x_i)\|_Y.$$

The objective inequality is obtained by

$$\begin{aligned} & P\left[w(\mathcal{F}_n, \Delta) \geq \epsilon_1\right] \\ \leq\ & P\left[\max_{1 \leq i \leq v} \sup_{s \in \bar{B}(x_i, 2\Delta)} \|\mathcal{F}_n(s) - \mathcal{F}_n(x_i)\|_Y \geq \frac{\epsilon_1}{2}\right] \\ \leq\ & \sum_{i=1}^{v} P\left[\sup_{s \in \bar{B}(x_i, 2\Delta)} \|\mathcal{F}_n(s) - \mathcal{F}_n(x_i)\|_Y \geq \frac{\epsilon_1}{2}\right]. \end{aligned}$$

By the Lipschitz condition imposed on φ_j, the condition

$$P\left(\left|\hat{\beta}^*_{nj}(\tilde{x}) - \hat{\beta}^*_{nj}(\tilde{x}^*)\right| \leq L_2 \|\tilde{x} - \tilde{x}^*\|_{X^N}, \tilde{x}, \tilde{x}^* \in K^N\right) = 1$$

and the condition for the number of points in the net for K^{N+1}, there exist an n_0^* and a Δ such that for $n \geq n_0^*$,

$$
\begin{aligned}
& P\left[w(\mathcal{F}_n, \Delta) \geq \epsilon_1\right] \\
\leq & \sum_{i=1}^{v} P\left[\sup_{s \in \bar{B}(x_i, 2\Delta)} \|\mathcal{F}_n(s) - \mathcal{F}_n(x_i)\|_Y \geq \frac{\epsilon_1}{2}\right] \\
\leq & \sum_{i=1}^{v} P\left\{\sum_{j=1}^{m}\left[\sup_{s \in \bar{B}(x_i, 2\Delta)} \left|\left|\hat{\beta}_{nj}^*(\tilde{s})\varphi_j(\breve{s}) - \hat{\beta}_{nj}^*(\tilde{x}_i)\varphi_j(\breve{x}_i)\right|\right|_Y\right] \geq \frac{\epsilon_1}{2}\right\} \\
\leq & \sum_{i=1}^{v} P\left\{\sum_{j=1}^{m}\left[\sup_{s \in \bar{B}(x_i, 2\Delta)} \left|\hat{\beta}_{nj}^*(\tilde{x}_i)\right| ||\varphi_j(\breve{s}) - \varphi_j(\breve{x}_i)||_Y\right] \geq \frac{\epsilon_1}{4}\right\} \\
& + \sum_{i=1}^{v} P\left\{\sum_{j=1}^{m}\left[\sup_{s \in \bar{B}(x_i, 2\Delta)} \left|\hat{\beta}_{nj}^*(\tilde{s}) - \hat{\beta}_{nj}^*(\tilde{x}_i)\right| ||\varphi_j(\breve{s})||_Y\right] \geq \frac{\epsilon_1}{4}\right\} \\
\leq & \sum_{i=1}^{v} P\left\{\sum_{j=1}^{m}\left|\hat{\beta}_{nj}^*(\tilde{x}_i)\right| \geq \frac{\epsilon_1}{8L_1\Delta}\right\} \\
& + \sum_{i=1}^{v} P\left\{\sum_{j=1}^{m}\sup_{s \in \bar{B}(x_i, 2\Delta)} \left|\hat{\beta}_{nj}^*(\tilde{s}) - \hat{\beta}_{nj}^*(\tilde{x}_i)\right| \geq \frac{\epsilon_1}{4M}\right\} \\
= & \sum_{i=1}^{v} P\left\{\sum_{j=1}^{m}\left|\hat{\beta}_{nj}^*(\tilde{x}_i)\right| \geq \frac{\epsilon_1}{8L_1\Delta}\right\} \\
\leq & v\left[P\left(z_{(m)}^* \geq \frac{\epsilon_1}{8mL_1\Delta}\right) + \epsilon_3\right] \\
\leq & \frac{2^{7/2}(2^m - 1)mL_1 h(\Delta^{-1})}{\pi^{1/2}\epsilon_1} \exp\left(\frac{-\epsilon_1^2}{128m^2L_1^2\Delta^2}\right) + v\epsilon_3 \\
\leq & \epsilon_2,
\end{aligned}
$$

for given positive numbers ϵ_1 and ϵ_2, where $M = \max_{1 \leq j \leq m} \sup_{\breve{x} \in K} \varphi_j(\breve{x})$, ϵ_3 associated with ϵ_2 depends on n, $z_{(m)}^*$ is the maximum of $|z_1|, \ldots, |z_m|$, the equality is due to

$$
P\left\{\sum_{j=1}^{m}\sup_{s \in \bar{B}(x_i, 2\Delta)} \left|\hat{\beta}_{nj}^*(\tilde{s}) - \hat{\beta}_{nj}^*(\tilde{x}_i)\right| < \frac{\epsilon_1}{4M}\right\} = 1
$$

as $\Delta < \epsilon_1/(8mML_2)$, and the fifth and the sixth inequalities are due to the mapping theorem and the following result,

$$\begin{aligned} & P\left(z^*_{(m)} \geq \frac{\epsilon_1}{8mL_1\Delta}\right) \\ = \; & 1 - \left[P\left(|z_j| < \frac{\epsilon_1}{8mL_1\Delta}\right)\right]^m \\ = \; & \sum_{j=1}^{m}(-1)^{j+1}2^j\binom{m}{j}\left[P\left(z_j \geq \frac{\epsilon_1}{8mL_1\Delta}\right)\right]^j \\ \leq \; & 2(2^m-1)P\left(z_j \geq \frac{\epsilon_1}{8mL_1\Delta}\right) \\ \leq \; & \frac{2^{7/2}(2^m-1)mL_1\Delta}{\pi^{1/2}\epsilon_1}\exp\left(\frac{-\epsilon_1^2}{128m^2L_1^2\Delta^2}\right). \end{aligned}$$

Note that the last inequality given in the above holds by employing the inequality for the tail of the standard normal distribution (Billingsley, 1999, M25). Finally, the result follows by Theorem 3.2.

♢

Remark 3.1. *Although the space $C(K^{N+1}, Y)$ might not be reflexive, it might be the subset of some reflexive Banach spaces endowed with another norm. For example, $C([0,1], R)$ is the subset of the Hilbert space $L_2([0,1])$ endowed with the inner product $< F_1, F_2 >_{L_2([0,1])} = \int_{[0,1]} F_1(x)F_2(x)dx$ and the norm induced by the inner product, i.e., the space of square-integrable functions on $[0,1]$ with respect to the Lebesgue measure, where F_1, F_2 are any elements in the Hilbert space. Therefore, if the space $C(K^{N+1}, Y)$ is the subset of some reflexive Banach space and the Schauder basis falls in the space $C(K^{N+1}, Y)$, the operator $\hat{F}_m$ corresponding to the subset S of the reflexive Banach space falls in $C(K^{N+1}, Y)$. In addition, the estimator of $\hat{F}_m$ based on the estimated coefficients is a $C(K^{N+1}, Y)$-valued random variable. Therefore, in such case, Theorem 2.1 states that the minimizer in the reflexive Banach space or the associated objective functional defined on the space is approximatable by the ones in the non-reflexive space $C(K^{N+1}, Y)$. On the other hand, the only requirement imposed on φ_j in Theorem 3.3 is the Lipschitz condition. That is, the result holds for any finite number of functions φ_j satisfying the Lipschitz condition. Therefore, as V is not a reflexive Banach space and Theorem 2.1 might not be true, Theorem 3.3 may still be true.*

4. Statistical Applications

4.1 Nonparametric Regression Models

Let

$$y = F(x) + \epsilon,$$

where F is usually a real-valued function on some subset Ω of R^q, for example, the compact subset of R^q, $x \in R^q$ is the regressor, and both y and ϵ are real-valued random variables.

For the above models, F can be assumed to be a real-valued function falling in $V = L_p(\Omega)$ with some sensible measures (Adams & Fournier, 2003, Chapter 2; Aliprantis & Border, 2006, Chapter 13), $1 < p < \infty$, i.e., $|F|^p$ being integrable and V being a separable and reflexive Banach space. For instance, if F is assumed to be square-integrable with respect to the Lebesgue measure on R^q and the integrated square error, i.e., $\int_\Omega [F(x) - F_0(x)]^2 dx = ||F - F_0||^2_{L_2(\Omega)}$, is employed, the existence of the unique minimizer and its approximatability can be proved by Corollary 2.1 (b), where F_0 is some specific function of interest, for example, the true function. Further, if the function $\phi = h^+$, $h^+(t) = t^2/2$ as $0 \leq t \leq M$, and $h^+(t) = Mt - M^2/2$ as $t > M$, $M > 0$, i.e., h^+ being the restriction of the Huber's function h (see Bauschke & Combettes, 2011, Example 8.35) to R^+, is employed, the existence of the robust minimizer of the objective function $h^+(||F - F_0||_{L_2(\Omega)})$ and the approximatability of the robust objective function can be proved by Corollary 2.1 (a). If the objective function is the residual sum of squares $\sum_{i=1}^n e_i^2$ in multivariate nonparametric regression models, F is assumed to fall in a separable reproducing kernel Hilbert space V with a reproducing kernel $R(\cdot, \cdot)$, and S is bounded, for example, $S = \{F : ||F||_V \leq r\}$ being a ball, $r > 0$, or S consisting of positive functions falling in the ball, theoretical results concerning the existence of the minimizer and the approximatability of the objective function can be proved by Corollary 2.3 (a), where (x_i, y_i) are the observations and $e_i = y_i - F(x_i)$. In addition, the existence of the robust minimizer and the approximatability of the objective function based on $\phi_i[F(x_i)] = h(e_i)$, i.e., the objective function being $\sum_{i=1}^n h(e_i)$, can be also proved by Corollary 2.3 (a). In some cases, the assumptions imposed on S can be relaxed. For some examples, if S is any set of which orthogonal projection on V_2 is a nonempty bounded closed convex set and the objective function is equal to $\sum_{i=1}^n e_i^2 + c\|P_{V_1^\perp}(F)\|_V^2$ (Wahba, 1990, Chapter 1) or $\sum_{i=1}^n h(e_i) + c\|P_{V_1^\perp}(F)\|_V^2$, i.e., the addition of the penalty term $c\|P_{V_1^\perp}(F)\|_V^2$, theoretical results concerning the existence of the nonrobust and robust minimizers, both linear combinations of a finite number of basis functions, can be proved by using Lemma 2.1, where V_2 is the space spanned by $\{R(, x_i), i = 1, \ldots, n\}$ and the basis functions in V_1, $c > 0$, $V_1^\perp$ is the orthogonal complement of any finite-dimensional subspace V_1 of V, and $P_{V_1^\perp}$ is the projection operator of V onto $V_1^{\perp}$. One example of the above set S is the set $\{F : F = F_1 + F_2, F_1 \in S_2, F_2 \in S_2^{\perp}\}$, where S_2 is any nonempty bounded closed convex subset of V_2, $S_2^{\perp}$ is any subset of $V_2^{\perp}$, and where $V_2^{\perp}$ is the orthogonal complement of V_2. Furthermore, as the resulting estimator is a linear combination of a finite number of continuous basis functions defined on some compact subset of R^q, both the consistency and weak convergence of the resulting estimator in the nonparametric regression models can be proved by Corollary 3.1 and Theorem 3.3.

4.2 Measurement Error in Nonparametric Regression Models

Let

$$y = F(x) + \epsilon$$

where x is a R^q-valued random vector and ϵ is a real-valued random variable.

For the above models, assume that X is the space consisting of random vectors with means and Y is the space consisting of random variables with finite variances, i.e., X being a separable Banach spaces and Y being a separable Hilbert space with the inner product $< y_1, y_2 >_Y = E(y_1 y_2)$ for $y_1, y_2 \in Y$. Further, assume that ϵ has a finite variance. If the inner product for the separable Hilbert space V is $< F_1, F_2 >_V = \sum_{i=1}^{n} < F_1(x_i), F_2(x_i) >_Y$ and the objective functions are the expected value of the error sum of squares $E(\sum_{i=1}^{n} \epsilon_i^2) = ||F - F_0||_V^2$ and $h^+(||F - F_0||_V)$, the existence of the unique nonrobust minimizer and the robust minimizer(s) and the associated approximatability can be proved by Corollary 2.1 (b) and Corollary 2.1 (a), respectively, where $x_i \in X$, $y_i \in Y$, $\epsilon_i = y_i - F(x_i)$, and $F_0(x_i) = y_i$. Further, if the penalized objective function $E(\sum_{i=1}^{n} \epsilon_i^2) + c\|P_{V_1^{\perp}}(F - F_0)\|_V^2$ is employed and $S - F_0 = \{F - F_0 : F \in S\}$, the shift of the set S, is a set of which orthogonal projection on V_1 is a nonempty bounded closed convex set, that the existence of the minimizer, a linear combination of a finite number of basis functions, can be proved by using Lemma 2.1, were V_1 is any finite dimensional subspace of V. One example of S is the set $\{F : F = F_0 + F_1 + F_2, F_1 \in S_1, F_2 \in S_1^{\perp}\}$, where S_1 is any nonempty bounded closed convex subset of V_1 and $S_1^{\perp}$ is any subset of $V_1^{\perp}$. In addition, both the consistency and weak convergence of the resulting estimator based on finite dimensional approximation in the Polish space consisting of continuous nonlinear operators on the compact subset of X (Aliprantis & Border, 2006, Chapter 3.19; Remark 3.1), i.e., a linear combination of a finite number of basis functions in the space, can be proved by Corollary 3.1 and Theorem 3.3.

Note that the other model to fit the data with the random regressor is

$$y = F[E(x)] + \epsilon.$$

For the model, theoretical results can be proved analogous to the ones in Section 4.1.

4.3 Functional Data Analysis

Let

$$y_j(x) = F_j(x) + \epsilon_{xj}, j = 1, \ldots, m$$

(Ramsay & Silverman, 2005), where $x \in R$, $F_j(x)$ are real-valued functions, and ϵ_{xj} are real-valued random variables.

The above model can be considered as the nonparametric regression models with m functions (curves). F_j can be assumed to be real-valued functions falling in a separable and reflexive Banach space V. Then, theoretical results analogous to the ones in Section 4.1 concerning the existence, the approximatability, the consistency, and weak convergence for the minimizers of interest and the associated objective functions can be proved based on the ones in Section 2 and Section 3.

4.4 Nonparametric Regression Models Using Differential Equations

Ordinary differential equations have been used for the fitting of the functional data (Ramsay & Silverman, 2005, Chapter 18 and Chapter 19). It might be also sensible to consider the general differential equations for data fitting, i.e., the partial differential equations (PDE). A dth order PDE is

$$D\left[x, F(x), \frac{\partial F(x)}{\partial x_1}, \ldots, \frac{\partial^{|\alpha|} F(x)}{\partial x_{j_1}^{\alpha_1} \ldots \partial x_{j_r}^{\alpha_r}}, \ldots, \frac{\partial^d F(x)}{\partial x_q^d}\right] = 0,$$

where $x = (x_1, \ldots, x_q)^t \in R^q$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r), |\boldsymbol{\alpha}| = \sum_{k=1}^r \alpha_k \leq d, \alpha_k$ are non-negative integers, $\{j_1, \ldots, j_r\} \subset \{1, \ldots, q\}$, F is a real-valued function, and D is a function defined on some subset of R^s. For example, a second order PDE with q=2 and s=8 is

$$D\left[x, F(x), \frac{\partial F(x)}{\partial x_1}, \frac{\partial F(x)}{\partial x_2}, \frac{\partial^2 F(x)}{\partial x_1^2}, \frac{\partial^2 F(x)}{\partial x_1 \partial x_2}, \frac{\partial^2 F(x)}{\partial x_2 \partial x_1}, \frac{\partial^2 F(x)}{\partial x_2^2}\right] = 0.$$

The weak solutions of several well-known PDEs subject to prescribed boundary and initial conditions fall in the Hilbert space, including the ones of second order elliptic PDEs, second order parabolic PDEs, second order hyperbolic PDEs, and Euler-Lagrange equation (Evans, 1998). Further, the weak solutions for the second order elliptic PDEs, the second order parabolic PDEs and the second order hyperbolic PDEs can be imbedded into the space of smooth functions, i.e., the subspace of some separable Hilbert space, while the weak solutions for the Euler-Lagrange equation can fall in a separable Hilbert space by imposing a few strong assumptions (Evans, 1998, Chapter 8.3). Therefore, it is sensible to assume that the solution F, possibly not unique, exists and falls in a separable Hilbert space.

Suppose that the model for the response y is

$$y = F(x) + \epsilon,$$

where ϵ is a real-valued random variable. If the total sum of squares given in Theorem 2.1 of Wei (2014) excluding the penalty term, i.e., the sum of those mean sum of squares corresponding to the responses, the differential equation, the initial conditions, and the boundary conditions, is employed and S is any set of which orthogonal projection on V_1 is a nonempty bounded closed convex set, i.e., the assumptions on S being relaxed, the existence of the minimizer and the approximatability of the objective function can be proved by using Theorem 2.1, where V_1 is the space spanned by the representers. For example, S can be the set $\{F : F = F_1 + F_2, F_1 \in S_1, F_2 \in S_1^{\perp}\}$, where S_1 is any nonempty bounded closed convex subset of V_1 and $S_1^{\perp}$ is any subset of $V_1^{\perp}$. In addition, if the resulting estimated solution is a linear combination of a finite number of continuous basis functions defined on some compact subset of R^q, both the consistency and weak convergence of the resulting estimator can be proved by Corollary 3.1 and Theorem 3.3.

References

Adams, R. A. and Fournier, J. J. F. (2003). *Sobolev Spaces*, 2nd ed. Academic Press, Boston.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. Springer-Verlag, Berlin Heidelberg.

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, **68**, 337-404.

Bauschke, H. H. and Combettes, P. L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, Springer, New York.

Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers, Boston.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

Da Prato G. and Zabczyk J. (1992). *Stochastic Equations in Infinite Dimensions*, Cambridge University Press, Cambridge.

Deimling, K. (1985). *Nonlinear Functional Analysis*, Springer-Verlag, Berlin Heidelberg.

Ekeland, I. and Témam, R. (1999). *Convex Analysis and Variational Problems*, SIAM, Philadelphia.

Enflo, P. (1973). A counterexample to the approximation problem in Banach spaces. *Acta Math.*, **130**, 309-317.

Evans, L. C. (1998). *Partial Differential Equations*, AMS, Providence.

Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*, Wiley, New York.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces*, Springer-Verlag, Berlin Heidelberg.

Meise, R. and Vogt, D. (1997). *Introduction to Functional Analysis*, Oxford University Press, New York.

Morrison, T. J. (2001). *Functional Analysis: An Introduction to Banach Space Theory*, Wiley, New York.

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

Seber, G. A. F. (1977). *Linear Regression Analysis*, Wiley, New York.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia.

Wei, W. H. (2014). Mathematical equation models. *J. Chinese Statist. Assoc.*, **52**, 497-532.

Wei, W. H. (2016). Weak convergence of probability measures on metric spaces of nonlinear operators. *Bull. Inst. Math. Acad. Sin. (N.S.)*, in press.

[Received January 2016; accepted July 2016.]

Journal of the Chinese
Statistical Association
Vol. 54, (2016) 186–204

無母數迴歸之有限維空間逼近論

魏文翔
東海大學統計系

摘　要

本文建立針對無母數迴歸 (nonparametric regression) 中被估計函數爲一定義及取值皆在可分離之實數場巴拿赫空間 (real separable Banach space) 的非線性算子 (nonlinear operator) 而其估計算子所形成空間爲有限維情形下發展相關理論結果。在估計過程中, 延伸出一新的概念, 稱之爲可逼近性 (approximatability)。被估計算子在不同狀況下是可逼近的理論結果被證明, 而估計算子之一致性 (consistency) 及弱收斂性 (weak convergence) 亦被證明。此外, 這些理論能應用到不同的統計模式。

關鍵詞: 可逼近性, 巴拿赫空間 (Banach space), 一致性, 非線性算子, 無母數迴歸, 弱收斂性。
JEL classification: C14, C61.