Tatra Mt. Math. Publ. **17** (1999), 227–239

Tatra Mountains
Mathematical Publications

NONPARAMETRIC REGRESSION ESTIMATION BY BERNSTEIN-DURRMEYER POLYNOMIALS

Ewaryst Rafajłowicz — Skubalska-Rafajłowicz Ewa

ABSTRACT. In the paper a new nonparametric estimator (based on Bernstein-Durrmeyer polynomials) of a regression function is proposed. The estimator is able to attain the best possible convergence rate in the MISF sense without any additional corrections of boundary effects. The extension to the multivariate case (using space-filling curves) is also discussed. Simulations indicate that the estimator is not too sensitive to an incorrect choice of the polynomial degree, which is the smoothing parameter.

1. Introduction and problem statement

Consider the problem of nonparametric estimation of a function $f\colon X = [0,1] \to \mathbb{R}$ in the fixed design case (see [8], [9] for extensive bibliographies). Observations y_i, $i = 1,2,\dots,n$ are of the form:

$$y_i = f(x_i) + \varepsilon_i\,, \qquad i = 1,2,\dots,n, \tag{1}$$

where ε_i, $i = 1,2,\dots,n$, are zero mean, uncorrelated random variables with finite variances $\sigma_i^2 \le \sigma^2 < \infty$ $i = 1,2,\dots,n$. Observations are taken at prescribed design points x_i, $i = 1,2,\dots,n$, ($x_n = 1$), which depend on n, but we suppress this in the notation. Only certain general requirements on f, e.g., continuity, are imposed. Consider the following estimator of f

$$\hat{f}_n(x) = \sum_{i=1}^{n} y_i \cdot \int\limits_{X_i} R_N(x,y)\,dy\,, \tag{2}$$

2000 Mathematics Subject Classification: 62G05.
Keywords: nonparametric regression, Bernstein-Durrmeyer polynomials, space-filling curves.
This research was supported by the Council for Scientific Research of Polish Government.

where $X_i \stackrel{\text{def}}{=} [\kappa_{i-1}, \kappa_i)$, $\kappa_o \stackrel{\text{def}}{=} 0$, $\kappa_n \stackrel{\text{def}}{=} 1$ are chosen in such a way that $x_i \in X_i$ and $\bigcup_{i=1}^{n} X_i = X$, $X_i \cap X_j = \emptyset$, $i \neq j$, while R_N is defined by

$$R_N(x, y) = (N+1) \cdot \sum_{k=0}^{N} B_k^{(N)}(x) B_k^{(N)}(y) . \tag{3}$$

In (3), $B_k^{(N)}(x) = \binom{N}{k} x^k (1-x)^{N-k}$, $k = 0, 1, \ldots, N$ is kth Bernstein polynomial. Estimator (2) is based on a modification of Bernstein polynomials proposed by Durrmeyer [7], and investigated in [5], [2], among others.

Remark 1. To motivate [2] and to explain the difference between the Bernstein and the Bernstein-Durrmeyer operators, consider the following operators acting on a function f

$$\sum_{k=0}^{N} f\Big(\frac{k}{N+1}\Big) \cdot B_k^{(N)}(x) , \tag{4}$$

and

$$\sum_{k=0}^{N} \int_0^1 f(t) B_k^{(N)}(t)\, dt \cdot B_k^{(N)}(x) = \int_0^1 f(t) R_N(t, x)\, dt . \tag{5}$$

The first one is the classical Bernstein polynomial, which approximates a given function f, continuous in $[0,1]$. The second one is the Bernstein-Durrmeyer operator, which is defined for $f \in L_2[0,1]$. The difference between them is in the way of forming coefficients of a Bernstein polynomial for a given function f. It is visible that (2) can be interpreted as an empirical form of (5), where the integration with unknown f is replaced by its piecewise constant approximation based on observations y_i, $i = 1, 2, \ldots n$.

The rest of the paper is organized as follows. In Section 2, conditions for mean integrated square error (MISE) consistency of $\hat{f}_n$ and its convergence rate are investigated. Then, we discuss the multivariate case. Finally, the results of simulations are briefly reported.

Motivations behind adding one more nonparametric estimator of regression are the following

1) $\hat{f}_n$ is intermediate between kernel smoothers and methods based on orthogonal expansions.
2) Shape preserving properties of Bernstein-Durrmeyer polynomials are visible when observations are corrupted by random noise (see Fig. 3).
3) $\hat{f}_n$ is robust with respect to an imprecise choice of the smoothing parameter N (see Fig. 3), what allows to recommend this approach to non-expert users.

4) The approach proposed here allows to construct other estimators with similar features, using, e.g., Szasz-Mirakyan operators.

Remark 2. For $f \in C[0,\infty)$, the Szasz-Mirakyan operator is defined by

$$\sum_{k=0}^{\infty} f(k/n) p_k(nx), \quad p_k(t) \stackrel{\text{def}}{=} e^{-t} t^k / k!, \quad k = 0, 1, \ldots \tag{6}$$

while in its modified version the sum in (6) is truncated to N first terms (see [23] for the definition and earlier references). Replacing in (6) $f(k/n)$ by $\int_0^\infty f(t) p_k(nt)$ one obtains an operator, which is structured analogously to the Bernstein-Durrmeyer operator, what allows to consider it as a candidate for constructing an estimator similar to (2).

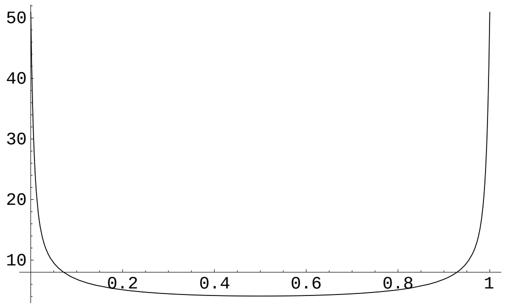

Figure 1. $R_N(x,x)$ for $N = 50$, $x \in (0,1)$.

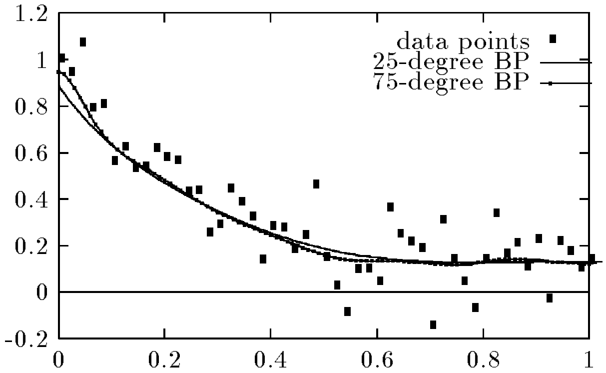

Figure 2.BDE fited to 50 observations from $f(x) + \exp(-5 \cdot x) + 0.1 + N(0, 0.1)$ noise.

Bernstein polynomials are popular in computer graphics (see [1], [12]). They were also used in [4] for nonparametric density estimation. However, we cannot apply these results directly in the fixed design regression estimation. In [22] a nonparametric estimator of a regression function has been proposed, which is based on the classical Bernstein polynomials. As we shall see, estimators based on the Bernstein-Durrmeyer polynomials have many remarkable features, which allow to prove the universal MISE consistency in L_2 and to assure the best possible convergence rate for the Lipschitz class of functions. Furthermore, this property holds without additional corrections of boundary effects, due to remarkable different behaviour of $R_N(x,y)$ in comparison to kernel estimates. This fact is illustrated in Fig. 1. Notice large values of $R_N(x,x)$ near the end points, while classical kernel smoothers are constant along the diagonal.

2. Regression estimation by Bernstein-Durrmeyer polynomials

Consider the equivalent form of (2)

$$\hat{f}_n(x) = (N+1) \cdot \sum_{k=0}^{N} \hat{b}_k \cdot B_k^{(N)}(x)\,, \tag{7}$$

$$\hat{b}_k = \sum_{i=1}^{n} y_i \int_{X_i} B_k^{(N)}(y)\,dy, \qquad k = 0, 1, \ldots, N\,. \tag{8}$$

Estimator (2) (and (7), (8)) will be further called the Bernstein-Durrmeyer estimator (BDE). Above, $N = N(n)$ is a sequence of nonnegative integers, which increases to infinity with the number of data points, in a way specified later.

The form (2) of $\hat{f}_n$ is similar to the estimate proposed by [3], while (7), (8) resemble the orthogonal series estimates discussed in [16], [14], and by Eubank [8]. However, properties of $\hat{f}_n$ are different from the properties of estimators mentioned above, due to lack of orthogonality of Berstein polynomials.

Our investigations of $\hat{f}_n$ are based on the following inequality

$$\int_X R_N(x,x)\,dx \le \sqrt{\pi \cdot (N+1)}/2\,. \tag{9}$$

Although this inequality seems to be new, we present its proof in Appendix B, since it is rather technical.

Lemma 1. (Variance). *Let* $\Delta_n \stackrel{\text{def}}{=} \max_{1 \le i \le n} |\kappa_i - \kappa_{i-1}|$. *Then,*

$$\operatorname{var}\big(\hat{f}_n(x)\big) \le (N+1)^2 \cdot \sum_{k=0}^{N} \operatorname{var}(\hat{b}_k) \cdot B_k^N(x)\,, \tag{10}$$

$$\int_X \operatorname{var}\big(\hat{f}_n(x)\big)\,,\ dx \le \sigma^2 \cdot \Delta_n \cdot \sqrt{\pi \cdot (N+1)}/2\,. \tag{11}$$

Proof. Applying the Schwarz inequality, we obtain from (7), (8):

$$\operatorname{var}(\hat{b}_k) \le \sigma^2 \cdot \Delta_n \cdot \int_X \big[B_k^{(N)}(x)\big]^2 dx\,, \tag{12}$$

$$\operatorname{var}(\hat{f}_n(x)) = (N+1)^2 \cdot E\left[\sum_{k=0}^{N} \hat{c}_k \cdot B_k^{(N)}(x)\right]^2, \tag{13}$$

where $\hat{c}_k \stackrel{\text{def}}{=} \hat{b}_k - E\hat{b}_k$. For fixed $x \in X$, $B_k^{(N)}(x)$ is a Bernouli probability distribution and the sum in the square brackets of (13) can be treated as the expectation with respect to this distribution. Furthermore, the quadratic function is convex, what allows to use the Jensen inequality to this expression. This implies (10). Now, (11) follows from (10) by integrating both of its sides and then using $\int\limits_X B_k^{(N)}(x)\,dx = (N+1)^{-1}$, $k = 0, 1, \ldots, N$, together with (12) and (9).

□

THEOREM 1. *Let $\hat{f}_n$ be given by (2) and let $N(n) \xrightarrow{n} \infty$ (as $n \to \infty$). Then for every $f \in L^2(X)$ the condition:*

$$\Delta_n \cdot \sqrt{N(n)} \xrightarrow{n} 0, \qquad \textit{implies} \quad E\|\hat{f}_n - f\|^2 \xrightarrow{n} 0\,. \tag{14}$$

Here and below $\|\hat{f}_n - f\|^2 \stackrel{\text{def}}{=} \int\limits_X \big(\hat{f}_n(x) - f(x)\big)^2 dx$ is the usual L_2 norm, while the expectation is taken with respect to distributions of all the random variables, which are present in the definition of $\hat{f}_n$.

P r o o f . It is clear that $E\big(\hat{f}_n(x) - f(x)\big)^2 = \operatorname{var}\big(\hat{f}_n(x)\big) + \big(f(x) - E\hat{f}_n(X)\big)^2$, and the integrated variance approaches zero by Lemma 1 (11) and (14). For the bias term we have $\big(f(x) - E\hat{f}_n(x)\big)^2 \le 2\cdot\big(f(x) - f_N(x)\big)^2 + 2\cdot\big(f_N(x) - E\hat{f}_n(x)\big)^2$, where $f_N \stackrel{\text{def}}{=} \mathbb{R}_N f \stackrel{\text{def}}{=} \int\limits_X R_N(x,y) f(y)\,dy$.

The first term approaches zero in $L_2(X)$ norm by Theorem **BP 1** given in Appendix A. For the second term

$$I_n(x) \stackrel{\text{def}}{=} \big|f_N(x) - E\hat{f}_n(x)\big| \le \sum_{k=0}^{N} \int\limits_X \big|f(y) - \bar{f}_n(y)\big| \cdot B_k^{(N)}(y)\,dy \cdot B_k^{(N)}(x)\,, \tag{15}$$

where $\bar{f}_n(x) \stackrel{\text{def}}{=} f(x_i)$ for $x \in [\kappa_{i-1}, \kappa_i)$, $i = 1, 2, \ldots, n$ From (15) we obtain: $\int\limits_X I_n^2(x)\,dx \le \big\|\mathbb{R}_N |f - \bar{f}_n|\big\|^2 \le \|\mathbb{R}_N\| \cdot \|f - \bar{f}_n\|^2$ and the last term approaches zero for every $f \in L^2(X)$, since $\|\mathbb{R}_N\| \le 1$ (see **BP 6** in Appendix A. □

MISE consistency of BDE holds without any further assumption on f other than the minimal requirement $f \in L_2(X)$. This property is not shared by many others nonparametric regression estimates. Next, we discuss the convergence rate of $\hat{f}_n$. Let $f \in \operatorname{Lip}(L, \alpha)$, then

$$E\|f - \hat{f}_n\|^2 \le \frac{\sqrt{\pi}}{2}\sigma^2 \Delta_n (N+1)^{1/2} + 2L^2 \Delta_n^{2\alpha} + 2\|f - \mathbb{R}_N f\|^2\,. \tag{16}$$

For two sequences α_n and β_n, $n = 1, 2, \ldots$, we write $\alpha_n \sim \beta_n$, if there exists two constants $c_1 > 0$ and $c_2 > 0$ such that $c_1 \le \alpha_n/\beta_n \le c_2$ for n sufficiently large.

COROLLARY 1. *Let* $N(n) \sim n^{\beta}$ *and* $\Delta_n \sim n^{-1}$. *If* $f \in \mathrm{Lip}(L, \alpha)$, $0 < \alpha \leq 1$, *then the best convergence rate of the r.h.s. of* (16) *is attained for* $\beta = 2/(1+2\cdot\alpha)$ *and* $E\|f - \hat{f}_n\|^2 = O\big(n^{-\frac{2\alpha}{1+2\alpha}}\big)$,

P r o o f. Follows directly from (16), taking into account that for $f \in \mathrm{Lip}(L, \alpha)$ we have $\varepsilon_N(f) = O(N^{-\alpha})$, (see Appendix A and [21]). □

It is known (see [8] and the bibliography there in) that the above convergence rate can not be exceeded by any other nonparametric estimator in the whole Lipschitz class.

3. Simulations

The aim of simulations was to gather an experience on the behaviour of different degrees BDE when fitted to data of a moderate size. Figures 2 and 3 present one sample behaviour (snap-shots) of BDE under various conditions.

Remark 3. We add that in practice $\hat{f}_n$ can be rescaled to any finite interval and applied to its subintervals. It is known (see [12]) that Bernstein polynomials defined in adjacent subintervals can be adjusted to ensure continuity of the resulting plot and its derivatives.

Simulations seem to confirm that BDE is easy to apply and reliable in recovering qualitative features of the estimated regression.

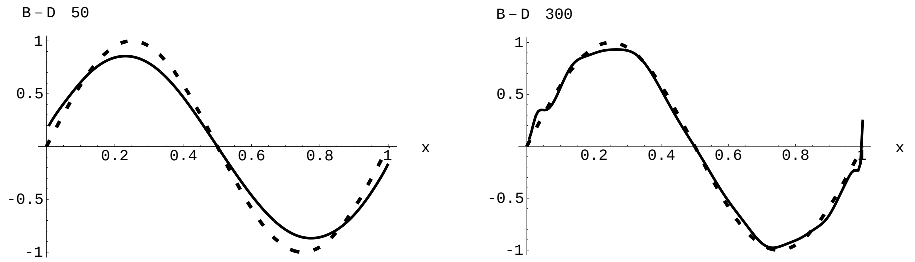

FIGURE 3. $\hat{f}_n$ fitted to $n = 50$ data simulated from $f(x) = \sin(2\pi x)$ (dashed line) + noise uniformly distributed in $[-0.1, 0.1]$. Plots for $N = 50, 300$ indicate that $\hat{f}_n$ is robust with respect to the choice of smoothing parameter N.

4. Extension to the multivariate case using space-filling curves

Extending estimator (2) to the multivariate case ($d \overset{\text{def}}{=} \dim(x) > 1$) by tensor products of the univariate Bernstein polynomials requires only elaboration of technical details in the proofs of convergence. We take another way of the generalization, which is based on scanning d-dimensional hypercube $I_d \stackrel{\text{def}}{=} [0,1]^d$ by a space-filling curve. This approach provides almost immediate proof of the existence of a multivariate Bernstein-Durrmeyer estimator, which attains the best possible MISE convergence rate in a multivariate case. Furthermore, one can obtain a rough estimate relatively quickly and using a relatively small number of observations, since — contrary to placing observations at grid points — one can place any desired number of points along a space-filling curve.

Let $\Phi\colon I_1 \to I_d$ be a space-filling curve, i.e., a transformation of $I_1 \stackrel{\text{def}}{=} [0,1]$ interval *onto* I_d. Well known examples of space-filling curves include the Peano and the Hilbert curves (see [17] and the bibliography there in). Sierpiński [18] elaborated a remarkable construction a closed two-dimensional curve. Multidimensional generalization of Sierpiński's curve was proposed by the second author in [19]. Any space-filling curve, for which the conditions (C1 – C5) given below hold, is suitable for our purposes.

C1) Φ is Lipschitz continuous, i.e., $\|\Phi(t_1) - \Phi(t_2)\| \le L_d |t_1 - t_2|^{1/d}$, $t_1, t_2 \in I_1$, where $\|.\|$ denotes the norm in I_d, $L_d > 0$ is a certain constant.
For $d \ge 2$, Φ is not differentiable. Φ is not invertible. We write $\Phi^{-1}(B) = \{t \in I_1\colon \Phi(t) \in B\}$ for $B \subset I_d$. Let μ_d be the Lebesque measure in I_d.

C2) Φ preserves the Lebesque measure.

C3) Φ is a.e. one-to-one and there exists $\Psi\colon I_d \to I_1$, being a.e. inverse of Φ.
Ψ is further called quasi-inverse of Φ.

C4) Ψ preserves the Lebesque measure, i.e., for every Borel set $A \subset I_1$ $\mu_1(A \cap \Psi(I_d)) = \mu_d(\Psi^{-1}(A \cap \Psi(I_d)))$.

C5) For every measurable function $g\colon I_d \to I_1$, $\int_{I_d} g(x)\,dx = \int_{I_1} g(\Phi(t))\,dt$.

In [11] or in [13], the results are formulated for the Peano, Hilbert and 2D Sierpiński curves, which imply that conditions C1) – C5) hold for them.

Now, we interpret f in (1) as a multivariate function $f\colon I_d \to \mathbb{R}$. For simplicity of exposition we assume that x_i's in (1) have the form $x_i = \Phi(t_i)$, $i = 1, 2, \dots n$, where t_i's are equidistantly placed in I_1 (see Fig. 4). Thus, $y_i = f(x_i) + \varepsilon_i = f(\Phi(t_i)) + \varepsilon_i$, $i = 1, 2, \dots, n$.

The idea of constructing estimator $\tilde{f}_n(x)$ of $f(x)$, $x \in I_d$ consists in using the correspondence between t_i and x_i, retaining the same observed value y_i, $i = 1, 2, \dots n$. Then, univariate estimator $\hat{f}_n$ is calculated from the sequence (t_i, y_i), $i = 1, 2 \dots, n$. Finally, its value at $t \in I_1$ is conveyed to point $I_d \ni x = \Phi(t)$. In other words, if $x \in I_d$ is the image of $t \in I_1$ by Φ, then $\tilde{f}_n(x)$ is defined as $\hat{f}_n(t)$, provided that $\hat{f}_n$ is calculated from data (t_i, y_i), $i = 1, 2 \dots, n$.

In more details,

$$\tilde{f}_n(x) = \sum_{i=1}^{n} y_i \cdot \int_{T_i} R_N\big(\Psi(x), \tau\big)\, d\tau\,, \tag{17}$$

where T_i's are subintervals of I_1, which form its partition and $t_i \in T_i$, $i = 1, 2, \ldots n$.

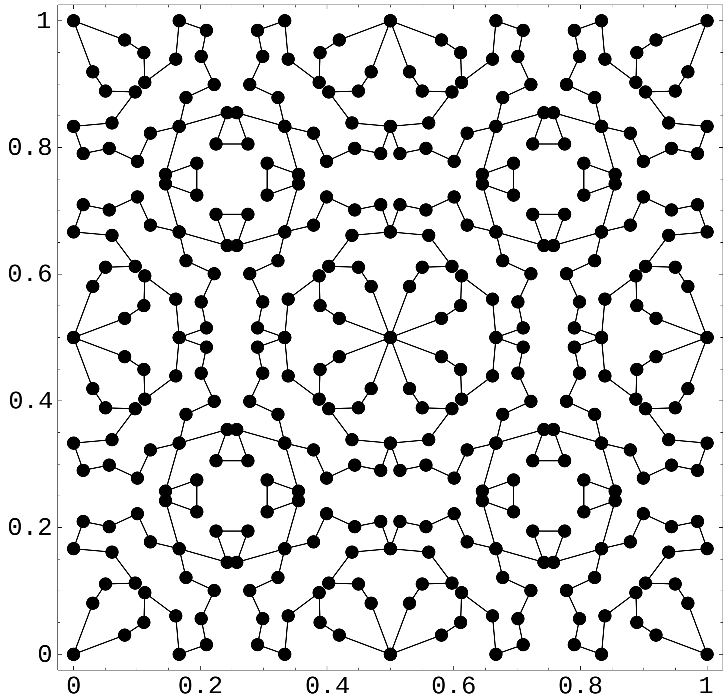

FIGURE 4. Points at which observations were taken, generated as 4×100 points along the Sierpiński curve. For each 100 points the Sierpiński curve was rescaled to fill one quarter of $[0,1] \times [0,1]$. For the visualisation purposes the points in each quarter of the unit square were joined by the corresponding Sierpiński curve.

THEOREM 2. *Let $\Delta_n = \max_{1 \le i \le n} \mu_1(T_i)$ and let $N = N(n) \to \infty$ an $n \to \infty$.*

1) *For every $f \in L^2(I_d)$, $\Delta_n \cdot \sqrt{N(n)} \xrightarrow{n} 0$ implies $E\|\tilde{f}_n - f\|^2 \xrightarrow{n} 0$.*
2) *Let f be Lipschitz continuous with the exponent $0 < \alpha \le 1$, i.e.,*

$$\big|f(x') - f(x'')\big| \le \|x' - x''\|^{\alpha}, \qquad x', x'' \in I_d\,. \tag{18}$$

Let $\Delta_n \sim n^{-1}$ and $N(n) \sim n^{\beta}$ with $\beta = \frac{2d}{d+2\cdot\alpha}$. Then $E\|f - \tilde{f}_n\|^2 = O\big(n^{-\frac{2\alpha}{d+2\alpha}}\big)$ and the exponent cannot be improved.

P r o o f . Form C5) we have

$$\int_{I_d} \big(\tilde{f}_n(x) - f(x)\big)^2 dx = \int_{I_1} \Big(\tilde{f}_n\big(\Phi(t)\big) - f\big(\Phi(t)\big)\Big)^2 dt\,. \tag{19}$$

From C2) it follows that $t = \Psi\big(\Phi(t)\big)$ a.e. in I_1. Thus, from (17) we have

$$\tilde{f}_n\big(\Phi(t)\big) = \sum_{i=1}^{n} y_i \cdot \int\limits_{T_i} R_N(t,\tau)\,d\tau\,. \tag{20}$$

Taking into account that $x_i = \Phi(t_i)$, we conclude that $\tilde{f}_n\big(\Phi(t)\big)$ coincides with $\hat{f}_n$ defined by (2), if the estimated function is given by $f\big(\Phi(t)\big)$, $t \in I_1$. Note also that by C5) $f(\cdot) \in L_2(I_d) \Longrightarrow f\big(\Phi(\cdot)\big) \in L_2(I_1)$. This fact, together with (19) and Thm. 1 completes the proof of Part 1. To prove Part 2, note that if for f condition (18) holds, then $f\big(\Phi(t)\big)$ is Lipschitz continuous but with the exponent α/d. Now, Part 2 follows from (19) and Corollary 1. Optimality of the rate $n^{-2\alpha/(d+2\alpha)}$ comes from Stone's result [20] (see also [8]). □

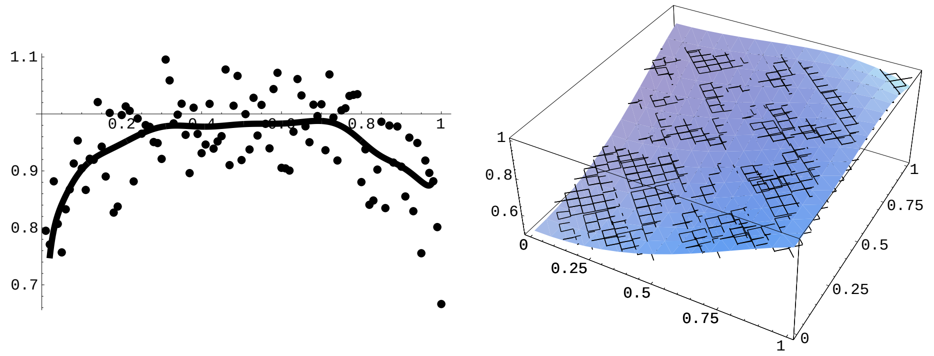

FIGURE 5. Left panel — univariate fit to $n = 100$ data in one of four subintervals mentioned in the text. Right panel — $\tilde{f}_n$ "glued" from four univariate fits (line) and "true" surface (shadowed). $\tilde{f}_n$ values were calculated at 2048 points along the Sierpiński curve. These points joined for the visualisation purposes only. Only a part of the curve estimating the surface is visible — the one above the "true" surface.

Analysis of the above proof indicates that the "curse of dimensionality" is hidden in the reduction of smoothness of $f\big(\Phi(t)\big)$ in comparison to f itself. For this reason, in practical applications of this approach it is useful to divide I_1 into several subintervals where $f(\Phi(t))$ is estimated, then to ensure continuity (see Remark 3) and to transform the values of the resulting estimate to I_d.

Results of applying $\tilde{f}_n$ to four subintervals of I_1 are shown in Fig. 5, where function $f\big(x^{(1)}, x^{(2)}\big) = \sin\Big(\pi\big((x^{(1)})^2 + (x^{(2)}\big)^2\Big)/3 + 1/2$ was estimated from $n = 100$ in each subinterval. These points were then transformed to the unit square using the space-filling curve (see Fig. 4). The observations were corrupted by errors uniformly distributed in $[-0.1, 0.1]$. In each subinterval $N = 50$ degree estimator was fitted.

Appendix A

The reader is referred to [10] for classical results on Bernstein polynomials (BP) and for historical account. We mention only the results necessary in this paper, which are taken from [6] [2], [5].

BP 1. 1) If $f \in L^2(X)$, then $\|f - \mathbb{R}_N f\| \to 0$ as $N \to \infty$.

2) Let $\varepsilon_N(f) \stackrel{\text{def}}{=} \inf\{\|f - w_N\|,\ w_N \in \Pi_N\}$, where Π_N is the class of all polynomials on X of order $\leq N$. Then, for $0 < \delta < 2$ and $N \to \infty$

$$\|f - \mathbb{R}_N f\| = O(N^{-\delta/2}) \qquad \text{iff} \quad \varepsilon_N(f) = O(N^{-\delta}).$$

BP 2. Kernel $R_N(x,y)$ is symmetric, degenerate and nonnegative definite, i.e., for every $g \in L^2(X)\langle \mathbb{R}_N g,\ g\rangle \geq 0$.

BP 3. (see [5]) Eigenfunctions l_k, $k = 0,1,\ldots,N$, of $\mathbb{R}_N$ are the orthonormal Legendre polynomials and the following expansion holds:

$$\mathbb{R}_N(x,y) = \sum_{k=0}^{N} \lambda_{kN} \cdot l_k(x) \cdot l_k(y), \tag{21}$$

$$\lambda_{0N} = 1 \quad \text{and} \quad \lambda_{kN} = \prod_{j=1}^{k} (N+1-j)/(N+1+j), \qquad k = 1,2,\ldots,N. \tag{22}$$

BP 4. For $k = 1,2,\ldots,N$, the following inequalities hold: $1 > \lambda_{1N} > \lambda_{2N} > \cdots > \lambda_{NN} > 0$ and $\lambda_{k(N+1)} > \lambda_{kN}$.

BP 5. For every $x, y \in X$, $R_N(x,y) \leq N+1$.

BP 6. The operator norm of $\mathbb{R}_N$ in $L^2(X)$ is bounded, i.e.,

$$\|\mathbb{R}_N\| = \sup_{\|f\|\leq 1} \langle f, \quad \mathbb{R}_N f\rangle \leq 1, \qquad N = 1,2,\ldots \tag{23}$$

To prove **BP 6**, it suffices to apply the Schwarz inequality to

$$\langle f, \mathbb{R}_N f\rangle = (N+1) \sum_{k=0}^{N} \left(\int_0^1 f(x) B_k^{(N)}(x)\, dx \right)^2,$$

then to note that $\left(B_k^{(N)}(x)\right)^2 \leq B_k^{(N)}(x)$ (since $B_k^{(N)}(x) \leq 1$) and finally to invoke the equality $\int_0^1 B_k^{(N)}(x) dx = 1/(N+1)$.

BP 7. For every $x \in [0,1]$, $\int_0^1 R_N(x,y)\, dy = 1$.

Appendix B

Proof of inequality (9)page 3. Our aim is to prove inequality (9), but firstly we need to prove some additional properties of the kernel $R_N(x, y)$.

BP 8. For $k = 0, 1, \ldots, N$ and $N \geq 1$

$$\lambda_{kN} \leq \exp\bigl[-k(k+1) \cdot (N+1)^{-1}\bigr]\,. \tag{24}$$

P r o o f o f **BP 8**. Using the inequality between the geometric and the arithmetic means we obtain from (22)

$$\lambda_{kN} = \prod_{j=1}^{k}\Bigl[1 - \frac{2j}{N+1+j}\Bigr] \leq$$
$$\Bigl[1 - \frac{2}{k} \cdot \sum_{j=1}^{k} \frac{j}{N+1+j}\Bigr]^k \leq \Bigl[1 - \frac{k+1}{N+1}\Bigr]^k. \tag{25}$$

Now, 8 follows from an elementary inequality

$$1 - \frac{t}{N+1} \leq \exp\bigl(-t/(N+1)\bigr) \text{ valid for } 0 \leq t \leq N+1\,.$$

□

Finally, we prove inequality (9), which is repeated here for convenience.

$$\int_X R_N(x, x)\, dx \leq \sqrt{\pi \cdot (N+1)}/2 \tag{26}$$

P r o o f o f (26). By 3, the left hand side of (26) equals to $\int_X R_N(x, x)\, dx = \sum_{k=0}^{N} \lambda_{kN}$. Applying 8 to this sum we get:

$$\int_X R_N(x, x)\, dx \leq \sum_{k=0}^{N} \exp\bigl[-k^2/(N+1)\bigr] \leq \int_0^\infty \exp\bigl[-y^2/(N+1)\bigr]\, dy\,, \tag{27}$$

which concludes the proof. □

Acknowledgements. The authors wish to express sincere thanks to an anonymous referee for useful comments resulting in more clear presentation of the paper.

REFERENCES

[1] BEZIER, P.: *Numerical Control. Mathematics and Applications*, Wiley and Sons, London, 1972.

[2] CHEN, W.—DITZIAN, Z.: *Best polynomial and Durrmeyer approximation in $l_p(s)$*, Indag. Math. **2** (1991), 437–452.

[3] CHENG, F. K.—LIN, P. E.: *Nonparametric estimation of a regression function*, Z. Wahrscheinlichkeitstheor. verw. Geb. **57** (1981), 223–233.

[4] CIESIELSKI, Z.: *Nonparametric polynomial density estimation*, Probab. Math. Statist. **9** (1988), 1–10.

[5] DERRIENIC, M. M.: *On multivariate approximation by Bernstein-type polynomials*, J. Approx. Theory **45** (1985), 155–166.

[6] DITZIAN, Z.—TOTIK, V.: *Moduli of Smoothness*, Springer-Verlag, Berlin, 1987.

[7] DURRMEYER, J. L.: *Une formule d'inversion de la transformée de Laplace: Application a la theorie des moments*, PhD thesis, These de 3e cycle, Faculté des Sciences de l'Université de Paris, 1967.

[8] EUBANK, R. L.: *Spline Smoothing and Nonparametric Regression*, Marcell Dekker, Inc., New York, Basel, 1988.

[9] HÄRDLE, W.: *Applied Nonparametric Regression*, Cambridge University Press, Boston, 1990.

[10] LORENTZ, G. G.: *Bernstein Polynomials*, University of Toronto Press, Toronto, 1953.

[11] MILNE, S. C.: *Peano Curves and Smoothness of Functions*, Adv. in Math., Vol. 35, pp. 129–157, 1980.

[12] PAVLIDIS, TH.: *Algorithms for Graphics and Image Processing*, Computer Science Press Inc., Maryland, 1982.

[13] PLATZMAN, L. K.—BARTHOLDI, J. J.: *Spacefilling Curves and the Planar Travelling Salesman Problem*, J. Assoc. Comput. Mach., Vol. 36, 1989, pp. 719–737.

[14] RAFAJLOWICZ, E.: *Nonparametric orthogonal series estimators of regression: a class attaining the optimal convergence rate in L_2*, Statist. and Probab. Lett. **15** (1987), pages 219–224.

[15] RAFAJLOWICZ, E.: *Nonparametric least squares estimation of a regression function*, Statistics **19** (1988), 349–358.

[16] RUTKOWSKI, L.: *On system identification by nonparametric function fitting*, IEEE Trans. Automat. Control **AC 27** (1982), 225–227.

[17] SAGAN, H.: *Space-filling curves*, Springer-Verlag, Berlin, Heidelberg, New York, 1994.

[18] SIERPIŃSKI, W.: *Sur une nouvelle courbe continue qui remplit toute une aire plane*, Bulletin de l'Acad. des Sciences de Cracovie A. (1912), 463–478.

[19] SKUBALSKA-RAFAJLOWICZ, E.: *The closed curve filling multidimensional cube*, Technical rep. **46/94**, ICT Technical University of Wrocław, (1994).

[20] STONE, C. J.: *Optimal global rate of convergence for nonparametric regression*, Ann. Statist. **10** (1982), 1040–1053.

[21] SZEGO, G.: *Orthogonal Polynomials, volume XXIII*, American Mathematical Society, New York, 1959.

[22] TENBUSCH, A.: *Nonparametric curve estimation with Bernstein estimates*, Metrika **45** (1997), 1–30.

[23] XIEHUA, S. : *On the convergence of the modified Szasz-Mirakjan operator*, Approx. Theory Appl. **10** (1994), 20–25.

Received March 17, 1998

Institute of Engineering Cybernetics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
P–Wrocław
POLAND

E-mail: raf@ldhpux.immt.pwr.wroc.pl