Acta Sci. Math., 47 (1948), 205—221

Linear combinations of iterated generalized Bernstein functions with an application to density estimation

WOLFGANG GAWRONSKI and ULRICH STADTMÜLLER

Starting from the classical theorem of Weierstraß (and its various modifications) on approximation of continuous functions by means of Bernstein polynomials (and its generalizations) in this paper a class of discrete and linear operators is developed. These operators are linear combinations of iterates of the original Bernstein type operators being constructed analogously to Fejér—Korovkin operators. Generalizing known results for the classical Bernstein case they approximate smooth functions more closely than the Bernstein type operators. Moreover, related operators for approximating derivatives are developed and these deterministic concepts are applied to probability density estimation for computing the mean square error of certain density estimators.

0. Introduction and summary

It is well known (e.g. [9]; [22]) that classical Bernstein polynomials and its various generalizations and modifications (such as e.g. generalized Bernstein polynomials of Szasz or Baskakov operators) approximate the associated function f with order $O(1/n)$ provided the derivative f' belongs to the class Lip 1. These operators are discrete, linear, and positive. More precisely, they are of form

$$B_n(f;\, x) := \sum_j f\left(\frac{j}{n}\right) p_{jn}(x); \quad f\in C(J), \tag{0.1}$$

where J throughout denotes one of the intervals $\mathbf{R}$, $[0, \infty]$, or $[0, 1]$ for simplicity, and the functions p_{jn} satisfy $p_{jn}(x)\geqq 0$, $x\in J$, $j\in \mathbf{Z}$, $n\in \mathbf{N}$. More generally, in this paper we admit $\{p_{jn}(x)\}_{j=-\infty}^{\infty}$ to be the n-fold convolution of a probability lattice distribution with expectation x (see Section 1) and we refer to (0.1) as a *generalized Bern-*

Received September 13, 1982.

stein function. Then the above order of approximation remains true for (0.1) (e.g. [1]; [2]; [3]; [9]; [12, Ch. 7]; [28]; see also Lemma 3).

In this paper we investigate quantities being related to or derived from the functions (0.1) thereby treating the following topics.

(i) An improvement of the rate of convergence provided f is sufficiently smooth (Section 2),

(ii) approximation of derivatives of f (Section 3), and as an application,

(iii) asymptotic of the mean square error (MSE) of an estimator for a probability density concentrated on J (Section 4).

(i) Dropping the positivity of the operator B_n we mention two methods for increasing the rate of convergence in case of classical Bernstein polynomials. One of them works by forming operators of type

$$(0.2)\qquad L_{r,n} := \sum_{i=1}^{r} a_{ri} B_{d_i n}, \quad 1 \leqq d_1 \leqq \ldots \leqq d_r, \quad a_{ri} \in \mathbf{R}$$

(e.g. [4]; [21], where more general singular integral operators with certain differentiability properties are discussed) whereas the second one uses the iterated Bernstein operator of Fejér—Korovkin type

$$(0.3)\qquad D_{r,n} := \sum_{i=1}^{r} \binom{r}{i} (-1)^{i-1} B_n^i$$

[11 and the references given there]. Both approximating functions $L_{r,n}(f; x)$ and $D_{r,n}(f; x)$ are polynomials the approximation order of which is $O(n^{-r})$ provided $f \in C_{2r}[0, 1]$. Besides the increase of the degree $L_{r,n}(f; x)$ has the disadvantage that f has to be evaluated at the points $j/d_i n$, $j=0, \ldots, n$, $i=1, \ldots, r$, whereas the use of $D_{r,n}(f; x)$ requires the knowledge of f at the distinct nodes $j/n, j=0, \ldots, n$ only. We use the second approach due to FELBECKER [11] and extend his result cited above to operators (0.3) based on (0.1) (Theorem 1). In particular this includes the classical Bernstein case treated in [11] which corresponds to $\{p_{jn}(x)\}$ as a binomial distribution and Szasz and Baskakov operators generated by Poisson's and the negative binomial distribution, respectively (see also Section 1).

(ii) If p_{jn}, as a function of $x \in J$, satisfies certain differentiability properties, then e.g. in [13], [30], [31] it was shown that the operators (0.1) are simultaneously approximating, i.e.

$$\left(\frac{d}{dx}\right)^s B_n(f; x) \to f^{(s)}(x),$$

$n \to \infty$, provided f fulfills certain smoothness and growth properties. However for higher derivatives the approximating functions become rather complicated expressions. Hence for approximating the s-th derivative of a function F on J we consider the discrete operators

(0.4) $$D_n^{(s)}(F;x) := n^s \sum_j p_{jn}(x)\Delta^s F\left(\frac{j}{n}\right),$$

where Δ is the forward difference operator acting on j. Then we prove a theorem on uniform approximation and a Voronowskaja property for $D_n^{(s)}$ (Theorems 2, 3).

(iii) If $s=1$, then motivated by (0.4) in [15], [16], [29] a smoothed histogram type estimator was developed for estimating an unknown probability density f concentrated on J. More generally, in Section 3 as an estimator for its r-th derivative we consider

(0.5) $$\hat{f}_N^{(r)}(x) := n^{r+1} \sum_j p_{jn}(x)\Delta^{r+1}\hat{F}_N\left(\frac{j}{n}\right),$$

where $\hat{F}_N$ denotes the empirical distribution function of an iid sample with density f. Extending the results in [15; 16; 29] we compute the exact order of magnitude for the MSE (Theorem 5)

$$E[(\hat{f}_N^{(r)}(x)-f^{(r)}(x))^2]$$

which turns out to be $\sim c\cdot N^{-4/(2r+5)}$ provided the scaling parameter n is chosen subject to $n=n(N)\sim N^{2/5}$, f is smooth enough and satisfies certain growth conditions. (Throughout $a_n\sim b_n$ means that $\lim_{n\to\infty} a_n/b_n=1$.) Dropping the property of positivity for an estimator of the density itself we construct an estimator suggested by the deterministic approximation operator (0.3). We replace (0.5) by $(r=0)$

(0.6) $$\hat{D}_{r,n}(x) := n \sum_j a_{jn}(x)\Delta\hat{F}_N\left(\frac{j}{n}\right)$$

as an estimator for $f(x)$, where $a_{jn}(x)$ depends on $p_{jn}(x)$ only. Then the order of the MSE of (0.6) is $N^{-4r/(4r+1)}$ when f is smooth enough again (Theorem 4). Comparable results for the most popular density estimator, the kernel estimator, give the same rate of mean square convergence [23].

In this paper we look at the deterministic approximation theorems from a probabilistic point of view, too. This is expressed by the technical treatment of the proofs where we use e.g. moment inequalities, Tschebyscheff's inequality, local central limit theorems and Edgeworth expansions of lattice distributions.

1. Auxiliary results

In this section we collect and prove some lemmata which are basic for the technical treatment of this paper. We suppose throughout that $\{p_{jn}(x)\}_{j=-\infty}^{\infty}$, $x\in J$, is the n-fold convolution of a lattice probability distribution $\{p_{j1}(x)\}$ concentrated on the integers and satisfying the following conditions:

$$p_{j1} \in C(J), \tag{1.1}$$

$$\sum_j p_{j1}(x) = 1, \quad \sum_j jp_{j1}(x) = x, \quad \sigma^2 = \sigma^2(x) := \sum_j (j-x)^2 p_{j1}(x), \tag{1.2}$$

$$|m|_k(x) := \sum_j |j-x|^k p_{j1}(x) < \infty \tag{M$_k$}$$

for some $k \in \mathbf{N}$, $k \geqq 2$, $x \in J$ and the convergence of the series is uniform on compact subsets of J. Further $\{p_{j1}(x)\}$ is assumed to have maximal span equal to 1, and if (M_k) holds, then we denote by

$$m_k(x) := \sum_j (j-x)^k p_{j1}(x)$$

the k-th central moment of $\{p_{j1}(x)\}$. For practical purposes obviously such lattice distributions are of interest for which the p_{j1} are "elementary" functions and the convolutions are easily computable in a closed form. Choices of particular interest are

(i) the binomial distribution

$$p_{jn}(x) = \binom{n}{j} x^j (1-x)^{n-j}, \quad 0 \leqq x \leqq 1,$$

(ii) Poisson's distribution

$$p_{jn}(x) = \frac{(nx)^j}{j!} e^{-nx}, \quad x \geqq 0,$$

(iii) the negative binomial distribution

$$p_{jn}(x) = \binom{n+j-1}{j} \frac{x^j}{(1+x)^{n+j}}, \quad x \geqq 0,$$

which produce for (0.1) Bernstein polynomials, Szasz and Baskakov operators, respectively. (See also the remarks at the end of Section 4.) Moreover, throughout U is a compact subinterval of J where $\sigma^2(x) \geqq \sigma_0^2 > 0$ holds.

Lemma 1. *Suppose that* (M_k), $k \geqq 2$, *holds, then*

(i) $\sum_j jp_{jn}(x) = nx$,

(ii) $\sum_j (j-nx)^2 p_{jn}(x) = n\sigma^2(x)$,

(iii) $\sum_j |j-nx|^k p_{jn}(x) \leqq A_k |m|_k(x) n^{k/2}$, *where* A_k *is a positive constant de-*

pending only on k,

(iv)*) $\sum_j (j-nx)^k p_{jn}(x) = k! \sum n(n-1)\dots(n-s+1) \prod_{r=1}^{k} \frac{1}{\nu_r!}\left\{\frac{m_r(x)}{r!}\right\}^{\nu_r} =$

$= \sum_{\nu=1}^{[k/2]} a_\nu(x)n^\nu$, *where the non-specified summation is taken over all integer solutions* $(\nu_1, \dots, \nu_k)$ *of the equations* $\nu_1+2\nu_2+\dots+k\nu_k=k$, $s=\nu_1+\dots+\nu_k$; *moreover, if* $k=2r$ *is even, then*

$$a_r(x) = \frac{(2r)!}{r!\,2^r}\sigma^{2r}(x).$$

Proof. (i), (ii) are trivial and (iii) is a form of Marcinkiewicz's inequality (e.g. [25, p. 41]); explicit bounds are given in [24, p. 60], [8]. The first equality in (iv) is obtained by using the k-th derivative of the characteristic function of $\{p_{jn}(x)\}$ computed via [24, Lemma 2, p. 135]. From the latter form the representation as polynomial in n is immediate. This polynomial has degree at most $[k/2]$, since $m_1(x)=0$, by (1.2) and because $\nu_1+2\nu_2+\dots+k\nu_k=k$, $\nu_1+\dots+\nu_k>k/2$ imply that $\nu_1\geqq 1$. Finally the form of $a_r(x)$ in case $k=2r$ is given in [7, Corollary 3 of Theorem 2, p. 294].

Using notations and properties of the difference operator in [12, p. 221] and a local central limit theorem [24, Theorem 17, p. 207, see also pp. 9, 139] in [14, Lemma 1] the following lemma is proved.

Lemma 2. (i) *Suppose that* (M_3) *holds,* $m\in N$, $\delta_n>0$ *and* $\delta_n\sqrt{n}\to\infty$. *Then for* $x\in U$ *we have*

$$\sum_{|j/n-x|\leqq\delta_n} p_{jn}(x)^m = \frac{1}{(2\pi\sigma^2(x)n)^{(m-1)/2}\sqrt{m}}(1+o(1)), \quad n\to\infty,$$

the o-term being independent of $x\in U$.

(ii) *If* (M_k) *holds with* r, $m\in N$, $k\geqq r+2$ *and* $\delta_n>0$ *with* $\delta_n\sqrt{n}\to\infty$ *as* $n\to\infty$, *then for* $x\in U$ *we have*

$$\sum_{|j/n-x|\leqq\delta_n} |\Delta^r p_{j-r,n}(x)|^m = \frac{\sigma\sqrt{n}\,c_{r,m}}{(\sqrt{2\pi}\,\sigma^{r+1}n^{(r+1)/2})^m}(1+o(1)), \quad n\to\infty.$$

Again the o-term is independent of $x\in U$ *and*

$$c_{r,m} := \int_{-\infty}^{\infty} |H_r(y)|^m e^{-my^2/2}\,dy,$$

where H_r *is the r-th Hermite polynomial defined e.g. in* [24, p. 139].

*) For $\xi\in\mathbf{R}$, $[\xi]$ denotes the largest integer not exceeding ξ, as customary.

Finally we mention an approximation theorem together with a Voronowskaja property for the operators B_n in (0.1) which is well known and has been treated in the literature in various modified versions (e.g. [1], [2], [3], [5], [12], [19], [20], [21], [28], [30], [31]) since it can be proved along standard lines we omit a proof.

Lemma 3. (i) *If* (M_k) *holds for some* $k\geqq 2$, *and if* $f\in C(J)$ *satisfies* $f(x)=O(x^k)$, $|x|\to\infty$, *then*

$$\lim_{n\to\infty} B_n(f;x)=f(x) \tag{1.3}$$

for all $x\in J$. *Moreover, the convergence is uniform on compact subsets of* J.

(ii) *If* (M_k) *holds for some* $k\geqq 3$, *and if* $f\in C_2(J)$ *satisfies* $f(x)=O(x^k)$, $|x|\to\infty$, *then*

$$\lim_{n\to\infty} n\{B_n(f;x)-f(x)\} = \frac{\sigma^2(x)}{2}f''(x) \tag{1.4}$$

for all $x\in J$. *Again the convergence is uniform on compact subsets of* J.

Considering instead of (0.1) the modification

$$B_n^*(f;x):=n\sum_j p_{jn}(x)\int_{j/n}^{(j+1)/n} f(y)\,dy \tag{1.5}$$

which can be looked at as generalized Kantorovich functions (cf. [5], [6], [20]) Lemma 3 remains true provided the right hand side of (1.4) is replaced by

$$B^*(f;x):=\frac{1}{2}(f'(x)+\sigma^2(x)f''(x)). \tag{1.6}$$

In the sequel for $f\in C_2(J)$ we use the notation

$$B(f;x):=\frac{\sigma^2(x)}{2}f''(x). \tag{1.7}$$

2. Iterated generalized Bernstein functions

In this section we treat topic (i) mentioned in the introduction. That is, generalizing Lemma 3 and extending partially the results in [11] we improve the rate of convergence in (1.3). Under (M_ν), $\nu\geqq 2$, we consider the growth condition

$$m_\nu(x)=O(x^\nu),\quad |x|\to\infty, \tag{2.1}$$

which in particular is satisfied for the examples cited in Section 1 and more generally, when $\sigma^2(x)$ is a polynomial of degree at most 2 and p_{j1} satisfies the differential equation $\sigma^2(x)p'_{j1}(x)=p_{j1}(x)(j-x)$ (cf. [13], [21], [30]). Then for functions f defined on J

and satisfying $f(x)=O(x^\nu)$, $|x|\to\infty$, in case of an unbounded interval J the *iterated generalized Bernstein functions*

$$B_n^r(f), \quad r\in\mathbf{N}_0, \tag{2.2}$$

are well defined and continuous on J. Here $B_n^0:=I$ is the identity operator, and $B_n^{r+1}(f):=B_n(B_n^r(f))$, $r\in\mathbf{N}_0$. Further we have by (2.1) and Lemma 1 (iii)

$$B_n^r(f;x) = O(x^\nu), \quad |x|\to\infty, \quad r\in\mathbf{N}_0, \tag{2.3}$$

the O-constant being independent of n. Moreover, $D_{k,n}$ (defined in (0.3)) can be written as

$$D_{k,n} = I-(I-B_n)^k, \quad k\in\mathbf{N}. \tag{2.4}$$

Then we prove

Theorem 1. *Suppose that* (M_k) *holds for some* $k\geqq 2r+1$, $r\in\mathbf{N}$, *and* m_ν *satisfies* (2.1) *and* $m_\nu\in C_{2r-2}(J)$ *for* $\nu=2,3,\ldots,2r$. *Moreover, assume that* $f\in C_{2r}(J)$ *and* $f^{(2r)}(x)=O(x^{k-2r})$, $|x|\to\infty$. *Then*

$$\lim_{n\to\infty} n^r\left(D_{r,n}(f;x)-f(x)\right) = (-1)^{r-1}B^r(f;x)$$

for all $x\in J$. *Further the convergence is uniform on compact subintervals of* J. *(The powers of* B *in* (1.7) *are inductively defined in the same way as those of* B_n.*)*

Proof. We proceed by induction with respect to r (see [11]). If $r=1$, then Theorem 1 is contained in Lemma 3 (ii). Hence we assume the assertion to be true for $r-1$, $r\geqq 2$. Let $K=[a,b]\subseteqq J$ and for $\delta>0$ we use the notation $K_\delta:=[a-\delta, b+\delta]$. Then we have

$$f(y) = \sum_{\nu=0}^{2r}(y-x)^\nu\frac{f^{(\nu)}(x)}{\nu!}+(y-x)^{2r}\varrho(y-x), \quad x\in K \tag{2.5}$$

where $\varrho(h)\to 0$, if $h\to 0$ and

$$(y-x)^{2r}\varrho(y-x) = O\left((y-x)^k\right), \quad |y|\to\infty \tag{2.6}$$

uniformly in $x\in K$. Now the Taylor series expansion (2.5) gives (use (1.2))

$$(B_n-I)(f;x) = \sum_{\nu=2}^{2r}\frac{f^{(\nu)}(x)}{\nu!}\sum_j\left(\frac{j}{n}-x\right)^\nu p_{jn}(x)+\sum_j\left(\frac{j}{n}-x\right)^{2r}\varrho\left(\frac{j}{n}-x\right)p_{jn}(x) =$$

$$=: \sum_{\nu=2}^{2r}\frac{1}{\nu!}\frac{f^{(\nu)}(x)}{n^\nu}\sum_j(j-nx)^\nu p_{jn}(x)+\xi_n(x).$$

Using (2.1), the differentiability properties of m_ν and the growth restriction on $f^{(2r)}$, by Lemma 1 (iv) we can write the latter identity as

$$(B_n-I)(f;x) = \sum_{s=1}^{2r-1}\frac{1}{n^s}g_s(x)+\xi_n(x),$$

where the functions g_s are independent of n and satisfy the conditions

$$g_s(x) = O(x^k), \quad |x| \to \infty, \tag{2.7}$$

and

$$g_s \in \begin{cases} C_{2(r-s)}(J), & 1 \leqq s \leqq r \\ C(J), & r < s \leqq 2r-1 \end{cases} \tag{2.8}$$

with $g_1(x) = B(f; x)$. Hence it follows that

$$\begin{aligned} n^r(B_n - I)^r(f; x) &= n^{r-1}(B_n - I)^{r-1}B(f; x) + \\ &+ \sum_{s=2}^{2r-1} n^{r-s}(B_n - I)^{r-1}(g_s; x) + (B_n - I)^{r-1}(n^r \xi_n; x). \end{aligned} \tag{2.9}$$

(Note that B_n can be applied to g_s, ξ_n, and $B(f)$, by (2.7), (2.6) and (2.3).)

Next, for functions $f_n \in C(J)$ satisfying

$$f_n(x) = O(x^k), \quad |x| \to \infty \tag{2.10}$$

uniformly in n we get (see (2.3); $M > 0$)

$$\begin{aligned} \sup_{x\in K} |(B_n - I)(f_n; x)| &\leqq \sup_{x\in K} |f_n(x)| + \sup_{x\in K} |B_n(f_n; x)| \leqq \\ \leqq \sup_{x\in K} |f_n(x)| + \sup_{x\in K} \left| \sum_{j/n\in K_\delta} f_n\left(\frac{j}{n}\right) p_{jn}(x) \right| &+ \sup_{x\in K} \left| \sum_{j/n\notin K_\delta} f_n\left(\frac{j}{n}\right) p_{jn}(x) \right| \leqq \\ \leqq 2 \sup_{x\in K_\delta \cap J} |f_n(x)| + M \sup_{x\in K} \sum_{j/n\notin K_\delta} &\left(\left|\frac{j}{n}\right|^k + 1 \right) p_{jn}(x) \leqq \\ \leqq 2 \sup_{x\in K_\delta\cap J} |f_n(x)| + M \sup_{x\in K} &\left(\left\{ \sum_{v=0}^{k} \binom{k}{v} \frac{|x|^{k-v}}{\delta^{k-v}} + \frac{1}{\delta^k} \right\} \sum_j \left| \frac{j}{n} - x \right|^k p_{jn}(x) \right) = \\ = 2 \sup_{x\in K_\delta\cap J} |f_n(x)| &+ O\left(\frac{1}{n^{k/2}}\right), \end{aligned}$$

by Lemma 1 (iii). Further, by (2.3) and (2.10) we may apply this estimate to $f_n(x) = (B_n - I)^{r-2}(f; x)$ and thus we obtain inductively (observe (2.7), (2.8))

$$\sup_{x\in K} |(B_n - I)^{r-1}(g_s; x)| \leqq$$

$$\leqq \begin{cases} 2^{s-2} \sup\limits_{x\in K_{(s-2)\delta}\cap J} |(B_n - I)^{r-s+1}(g_s; x)| + O\left(\frac{1}{n^{k/2}}\right), & 2 \leqq s \leqq r \\ 2^{r-1} \sup\limits_{x\in K_{(r-1)\delta}\cap J} |g_s(x)| + O\left(\frac{1}{n^{k/2}}\right), & r < s \leqq 2r-1 \end{cases}$$

as $n \to \infty$. Now from (2.9), by the induction hypothesis, we have

$$n^{r-1}(B_n - I)^{r-1}B(f; x) \to B^{r-1}(B(f; x)) = B^r(f; x)$$

and, since $k \geqq 2r+1$,

$$n^{r-s}(B_n - I)^{r-1}(g_s; x) \to 0, \quad s \geqq 2,$$

as $n \to \infty$, uniformly on K. Finally we conclude from (2.5), (2.6), and Lemma 1 (iii) that $(M, \varepsilon, \delta > 0)$

$$n^r |\xi_n(x)| \leqq \varepsilon + M n^{r-k} \sum_{|j-nx| > \delta n} |j-nx|^k p_{jn}(x) \leqq \varepsilon + \frac{M}{\sqrt{n}} \leqq 2\varepsilon$$

when n is large enough. This and (2.4) complete the proof.

In later applications (see Section 4) we need a modification of Theorem 1 for the generalized Kantorovich operators B_n^* (defined in (1.5)). Putting

$$D_{k,n}^* := \sum_{i=1}^{k} \binom{k}{i} (-1)^{i-1} B_n^{*i}, \quad k \in \mathbf{N}, \tag{0.3)*}$$

we have

$$D_{k,n}^* = I - (I - B_n^*)^k, \quad k \in \mathbf{N}. \tag{2.4)*}$$

Since the following theorem is proved along the same lines as the preceeding one we omit its proof and only state

Theorem 1*. *Under the assumptions of Theorem 1 we have*

$$\lim_{n \to \infty} n^r (D_{r,n}^*(f; x) - f(x)) = (-1)^{r-1} B^{*r}(f; x)$$

for all $x \in J$. Again the convergence is uniform on compact subsets of J.

3. Approximation of derivatives

In this section we treat topic (ii) mentioned in the introduction; that is, we prove an approximation theorem for the operators (0.4) together with a Voronowskaja property. In the sequel Δ denotes the difference operator defined by $\Delta a_j := a_{j+1} - a_j$ acting on a sequence $\{a_j\}$ (e.g. [12, p. 221]). For differences of higher order we have

$$\Delta^r a_j = \sum_{\nu=0}^{r} \binom{r}{\nu} (-1)^{r-\nu} a_{j+\nu}, \quad r \in \mathbf{N}_0, \tag{3.1}$$

where $\Delta^0 a_j := a_j$.

Theorem 2. *Suppose that* (M_k) *holds for some* $k \geqq 2$ *and* $F \in C_s(J)$, $s \geqq 0$, *satisfies* $F^{(s)}(x) = O(x^k)$, $|x| \to \infty$. *Then*

$$\lim_{n \to \infty} D_n^{(s)}(F; x) = F^{(s)}(x)$$

for all $x \in J$ and the convergence is uniform on compact subsets of J.

Proof. First we note that for $F\in C_s(\mathbf{R})$

$$\Delta^s F\left(\frac{j}{n}\right)=\frac{1}{n^s}F^{(s)}(\xi_{jn}) \tag{3.2}$$

with some $\xi_{jn}\in[j/n,(j+s)/n]$. Extending F suitably from J on $\mathbf{R}$ we have, by (3.2) $(\varepsilon,\delta>0)$

$$|D_n^{(s)}(F;x)-F^{(s)}(x)|=\left|\sum_j\left(F^{(s)}(\xi_{jn})-F^{(s)}(x)\right)p_{jn}(x)\right|\leqq$$

$$\leqq\varepsilon+\sum_{|\xi_{jn}-x|>\delta}|F^{(s)}(\xi_{jn})-F^{(s)}(x)|p_{jn}(x).$$

Further, restricting x to a compact subset of J with positive constants M, M' we get (see Lemma 1 (iii))

$$|D_n^{(s)}(F;x)-F^{(s)}(x)|\leqq\varepsilon+M\sum_{|\xi_{jn}-x|>\delta}|\xi_{jn}-x|^k p_{jn}(x)\leqq$$

$$\leqq\varepsilon+M\sum_j\left(\frac{s}{n}+\left|\frac{j}{n}-x\right|\right)^k p_{jn}(x)\leqq\varepsilon+\frac{M'}{n^{k/2}}\leqq 2\varepsilon$$

if n is large enough. This completes the proof.

The exact rate of convergence is given by the following Voronowskaja property.

Theorem 3. *Suppose that* (M_k) *holds for some* $k\geqq 3$ *and* $F\in C_{s+2}(J)$, $s\geqq 0$, *satisfies* $F^{(s+2)}(x)=O(x^{k-2})$, $|x|\to\infty$. *Then*

$$\lim_{n\to\infty} n\left(D_n^{(s)}(F;x)-F^{(s)}(x)\right)=\frac{1}{2}\left(sF^{(s+1)}(x)+\sigma^2(x)F^{(s+2)}(x)\right) \tag{3.3}$$

for all $x\in J$, *the convergence being uniform on compact subsets of* J.

Remark. If $s=0,1$, then the right hand side of (3.3) can be written as $B(F;x)$ and $B^*(F';x)$, respectively. This exhibits Theorem 3 as a generalization of Lemma 3 (ii) and the corresponding analogue for B_n^*. (See the remarks following Lemma 3.)

Proof of Theorem 3. Extending F suitably from J on $\mathbf{R}$ if necessary, since $F\in C_{s+2}(J)$, we have $(0\leqq\nu\leqq s)$

$$F\left(\frac{j+\nu}{n}\right)=\sum_{\mu=0}^{s+1}\frac{1}{\mu!}F^{(\mu)}\left(\frac{j}{n}\right)\left(\frac{\nu}{n}\right)^\mu+\frac{1}{(s+2)!}F^{(s+2)}(\xi_{j\nu})\left(\frac{\nu}{n}\right)^{s+2}$$

with $j/n\leqq\xi_{j\nu}\leqq(j+\nu)/n$ and further, by (3.1),

$$\Delta^s F\left(\frac{j}{n}\right)=\sum_{\mu=0}^{s+1}\frac{F^{(\mu)}(j/n)}{n^\mu\mu!}\sum_{\nu=0}^{s}\binom{s}{\nu}(-1)^{s-\nu}\nu^\mu+$$

$$+\frac{1}{n^{s+2}(s+2)!}\sum_{\nu=0}^{s}\binom{s}{\nu}(-1)^{s-\nu}\nu^{s+2}F^{(s+2)}(\xi_{j\nu}).$$

Since, by (3.1), for $a_j=j^\mu$ and $j=0$

$$\frac{1}{s!}\sum_{v=0}^{s}\binom{s}{v}(-1)^{s-v}v^\mu=\frac{1}{s!}\Delta^s j^\mu=\begin{cases}0, & 0\leqq\mu<s,\\ 1, & \mu=s,\\ \binom{s+1}{2}, & \mu=s+1,\end{cases}$$

we obtain

$$D_n^{(s)}(F;x)=\sum_j F^{(s)}\left(\frac{j}{n}\right)p_{jn}(x)+\frac{s}{2n}\sum_j F^{(s+1)}\left(\frac{j}{n}\right)p_{jn}(x)+$$

$$+\frac{1}{n^2(s+2)!}\sum_{v=0}^{s}\binom{s}{v}(-1)^{s-v}v^{s+2}\sum_j F^{(s+2)}(\xi_{jv})p_{jn}(x)=$$

$$=B_n(F^{(s)};x)+\frac{s}{2n}B_n(F^{(s+1)};x)+O\left(\frac{1}{n^2}\right)$$

the O-term being independent of x as long as x is restricted to a compact subset of J. Now Lemma 3 completes the proof.

4. MSE for density estimators of Bernstein type

In this section let f be an unknown probability density concentrated on a known interval J. Starting from the empirical distribution function $\hat{F}_N$ based on iid observations $X_1, \ldots, X_N$ having density f, Theorem 2 shows that, $\hat{f}_N^{(r)}(x)$, defined in (0.5), is an asymptotically unbiased estimator for the r-th derivative $f^{(r)}(x)$ provided f satisfies certain growth and smoothness conditions. If $r=0$, then for various particular cases in [15], [16], [29] the asymptotic of the MSE was computed. Asymptotic distributions for $\hat{f}_N^{(0)}$ have been derived in [27]. Based on Theorem 1* and Theorem 3 now we accelerate the mean square convergence, when $r=0$, and determine the asymptotic behaviour of the MSE for $\hat{f}_N^{(r)}(x)$ if $r\geqq 0$.

First dropping the positivity of $\hat{f}_N^{(0)}$ and motivated by Theorem 1* we consider

$$\hat{D}_{r,n}(x):=n\sum_j a_{jn}(x)\Delta\hat{F}_N\left(\frac{j}{n}\right) \tag{4.1}$$

with

$$a_{jn}(x):=\sum_{i=1}^{r}\binom{r}{i}(-1)^{i-1}B_n^{*i-1}(p_{jn};x) \tag{4.2}$$

as an estimator for $f(x)$.

Theorem 4. *Suppose that* (M_k) *holds for some* $k\geqq 2r+1$, $r\in\mathbf{N}$, m_v *satisfies* (2.1) *and* $m_v\in C_{2r-2}(J)$ *for* $v=2, 3, \ldots, 2r$. *Moreover assume that* $f\in C_{2r}(J)$ *and* $f^{(2r)}(x)=O(x^{k-2r})$, $|x|\to\infty$.

(i) *If* $\sigma^2(x)>0$, *then*

$$E((\hat{D}_{r,n}(x)-f(x))^2) = \left\{\frac{B^{*r}(f;x)}{n^r}\right\}^2 + o\left(\frac{1}{n^{2r}}\right) + \frac{1}{N}V_n(x) \quad as \quad n\to\infty,$$

where

$$V_n(x) = \frac{f(x)\sqrt{n}}{2\sqrt{\pi}\,\sigma(x)} + o(\sqrt{n}), \quad if \quad r=1$$

$$|V_n(x)| \leqq \frac{(2^r-1)^2}{\sqrt{2\pi}\,\sigma_0} f(x)\sqrt{n} + O(1), \quad if \quad r>1.$$

Here the remainder terms hold uniformly on $U\subseteqq J$ *and* $\sigma(x)\geqq\sigma_0$ *for* $x\in U$.

(ii) *If* $\sigma^2(x)=0$, *then as* $n\to\infty$

$$E((\hat{D}_{r,n}(x)-f(x))^2) = \left\{\frac{f^{(r)}(x)}{2^r n^r}\right\}^2 + o\left(\frac{1}{n^{2r}}\right) + \frac{f(x)n}{N} + \frac{1}{N}o(n).$$

Proof. We decompose

$$E((\hat{D}_{r,n}(x)-f(x))^2) = (E(\hat{D}_{r,n}(x))-f(x))^2 + \mathrm{Var}\,(\hat{D}_{r,n}(x))$$

as a sum of bias squared and a variance term. If F denotes the distribution function of f, then an application of Theorem 1* yields

$$(4.3)\qquad E(\hat{D}_{r,n}(x)) = n\sum_j a_{jn}(x)\Delta F\left(\frac{j}{n}\right) = \sum_{i=1}^{r}\binom{r}{i}(-1)^{i-1}B_n^{*i}(f;x) =$$

$$= D^*_{r,n}(f;x) = f(x) + \frac{(-1)^{r-1}}{n^r}B^{*r}(f;x) + o\left(\frac{1}{n^r}\right),$$

where the o-term holds uniformly on compact subsets of J. For the variance we note that

$$(4.4)\qquad \mathrm{Cov}\left(\Delta\hat{F}_N\left(\frac{j}{n}\right), \Delta\hat{F}_N\left(\frac{k}{n}\right)\right) = \frac{1}{N}\Delta F\left(\frac{j}{n}\right)\left(\delta_{jk} - \Delta F\left(\frac{k}{n}\right)\right)$$

and obtain

$$(4.5)\qquad \mathrm{Var}\,(\hat{D}_{r,n}(x)) = n^2\sum_{j,k}\mathrm{Cov}\left(\Delta\hat{F}_N\left(\frac{j}{n}\right), \Delta\hat{F}_N\left(\frac{k}{n}\right)\right)a_{jn}(x)a_{kn}(x) =$$

$$= \frac{n^2}{N}\left\{\sum_j a_{jn}(x)^2\Delta F\left(\frac{j}{n}\right) - \left(\sum_j a_{jn}(x)\Delta F\left(\frac{j}{n}\right)\right)^2\right\}.$$

(i) Suppose that $\sigma^2(x)>0$. If $r=1$, then $a_{jn}(x)=p_{jn}(x)$ and, by Lemma 2(i) it is easily shown that

$$(4.6)\qquad n^{3/2}\sum_j p_{jn}(x)^2\Delta F\left(\frac{j}{n}\right) = \frac{f(x)}{2\sqrt{\pi}\,\sigma(x)} + o(1), \quad n\to\infty,$$

uniformly on U. If $r \geqq 2$, then we use a local central limit theorem (see formula (1.4) in [14]; or Theorem 1 in [24, p. 207]) and obtain

$$B_n^{*i}(p_{jn};\, x) \leqq \frac{1}{\sqrt{2\pi n}\,\sigma_0} + O\left(\frac{1}{n}\right), \quad i \in \mathbf{N}_0,$$

where the O-term holds uniformly in $x \in U$. From this we get, by (4.2),

$$|a_{jn}(x)| \leqq \frac{2^r - 1}{\sqrt{2\pi n}\,\sigma_0} + O\left(\frac{1}{n}\right)$$

and thus, by the remarks following Lemma 3,

$$n^{3/2} \sum_j a_{jn}(x)^2 \Delta F\left(\frac{j}{n}\right) \leqq \left\{\frac{(2^r-1)}{\sqrt{2\pi}\,\sigma_0} + O\left(\frac{1}{\sqrt{n}}\right)\right\} \sum_{i=1}^{r} \binom{r}{i} B_n^{*i}(f;\, x) = \tag{4.7}$$

$$= \frac{(2^r-1)}{\sqrt{2\pi}\,\sigma_0} f(x) + O\left(\frac{1}{\sqrt{n}}\right)$$

uniformly on U. Now a combination of (4.3), (4.5), (4.6), (4.7) completes the proof of part (i).

(ii) In case $\sigma^2(x)=0$, we have $p_{jn}(x)=\delta_{j,nx}$ for some $j \in \mathbf{Z}$, δ_{jk} being Kronecker's symbol. This implies that $a_{jn}(x)=p_{jn}(x)=\delta_{j,nx}$, by (4.2). Using (1.6), (4.3) and (4.4) we find part (ii).

In case $\sigma^2(x)>0$ obviously the choice $n=n(N)\sim cN^{2/(4r+1)}$, $c>0$, $N\to\infty$, yields the estimate $(r>1)$

$$E(\hat{D}_{r,n}(x)-f(x))^2 = O(N^{-4r/(4r+1)}), \quad N\to\infty.$$

For corresponding kernel estimators (cf. [18, section 4]) $N^{-4r/(4r+1)}$ is the exact order of magnitude. By more careful estimates of $V_n(x)$ the constant involved in the leading term could be reduced.

Finally we extend Theorem 1 in [16] by

Theorem 5. *Suppose that* (M_k) *holds for some* $k \geqq \max(3, r+2)$, $r \in \mathbf{N}_0$. *Further, assume that* $f \in C_{r+2}(J)$ *and* $f^{(r+2)}(x)=O(x^{k-2})$, $|x|\to\infty$.

i) *If* $\sigma^2(x) > 0$, *then*

$$E((\hat{f}_N^{(r)}(x)-f^{(r)}(x))^2) = \left\{\frac{(r+1)f^{(r+1)}(x)+\sigma^2(x)f^{(r+2)}(x)}{2n}\right\}^2 + o\left(\frac{1}{n^2}\right) +$$

$$+\frac{f(x)c_{r,2}n^{r+1/2}}{2\pi\sigma^{2r+1}(x)N} + \frac{1}{N}o(n^{r+1/2}),$$

as $n\to\infty$, *where* $c_{r,2}$ *is defined in Lemma 2. Again the* o*-terms hold uniformly on* U.

(ii) *If* $\sigma^2(x)=0$, *then*

$$E((\hat f_N^{(r)}(x)-f^{(r)}(x))^2)=\left(\frac{f^{(r)}(x)}{2n}\right)^2+o\left(\frac{1}{n^2}\right)+\frac{f(x)4^r n^{2r+1}}{N}+\frac{1}{N}O(n^{2r}),\quad n\to\infty.$$

Proof. Due to the standard decomposition

$$E((\hat f_N^{(r)}(x)-f^{(r)}(x))^2)=(E(\hat f_N^{(r)}(x))-f^{(r)}(x))^2+\mathrm{Var}(\hat f_N^{(r)}(x))$$

we treat each summand separately. By Theorem 3 ($s=r+1$) and (0.5) we get ($F'(x)=f(x)$)

(4.8) $$E(\hat f_N^{(r)}(x))=D_n^{(r+1)}(F;x)=f^{(r)}(x)+$$

$$+\frac{1}{2n}((r+1)f^{(r+1)}(x)+\sigma^2(x)f^{(r+2)}(x))+o\left(\frac{1}{n}\right)$$

uniformly on compact subsets of J. For evaluating the variance we use partial summation (see also [14]) and obtain from (0.5)

$$\hat f_N^{(r)}(x)=(-1)^r n^{r+1}\sum_j \Delta^r p_{j-r,n}(x)\Delta\hat F_N\left(\frac{j}{n}\right).$$

Hence, by (4.4), we have

$$\mathrm{Var}(\hat f_N^{(r)}(x))=\frac{n^{2(r+1)}}{N}\sum_j(\Delta^r p_{j-r,n}(x))^2\Delta F\left(\frac{j}{n}\right)-\frac{1}{N}(E(\hat f_N^{(r)}(x)))^2=:\mathrm{I}-\mathrm{II},$$

say. For II we have by (4.8)

(4.9) $$\mathrm{II}=\frac{1}{N}O(1),\quad n\to\infty.$$

Next, by the continuity of f at $x\in U$ ($\varepsilon,\delta>0$) we get

$$\mathrm{I}-\frac{n^{2r+1}f(x)}{N}\sum_{|j/n-x|\le\delta}(\Delta^r p_{j-r,n}(x))^2=$$

$$=\frac{n^{2r+1}}{N}\sum_{|j/n-x|\le\delta}n\int_{j/n}^{(j+1)/n}(f(y)-f(x))\,dy\,(\Delta^r p_{j-r,n}(x))^2+$$

$$+\frac{n^{2(r+1)}}{N}\sum_{|j/n-x|>\delta}(\Delta^r p_{j-r,n}(x))^2\Delta F\left(\frac{j}{n}\right)=:\mathrm{I}'+\mathrm{II}',$$

say, and

(4.10) $$|\mathrm{I}'|\le\varepsilon\frac{n^{2r+1}}{N}\sum_{|j/n-x|\le\delta}(\Delta^r p_{j-r,n}(x))^2.$$

Writing ($\Delta F(j/n)\le 1$)

$$\mathrm{II}'\le\frac{n^{2(r+1)}}{N}\sum_{\nu,\mu=1}^{r}\binom{r}{\nu}\binom{r}{\mu}\sum_{|j/n-x|>\delta}p_{j+\nu-r,n}(x)p_{j+\mu-r,n}(x),$$

the use of Cauchy's inequality combined with Lemma 1 (iii) yields

$$\mathrm{II}' = \frac{1}{N} O(n^r), \quad n \to \infty.$$

Putting this together with (4.8)—(4.10), and Lemma 2 (ii) we have established part (i).

In case $\sigma^2(x)=0$, direct computation of I (note that $p_{jn}(x)=\delta_{j,nx}$ and x is an integer) yields

$$\mathrm{I} = \frac{n^{2r+2}}{N} \sum_j \Delta F\left(\frac{j}{n}\right)\left\{\sum_{\nu=0}^{r}\binom{r}{\nu}(-1)^{r-\nu}\delta_{j+\nu-r,nx}\right\}^2 =$$

$$= \frac{n^{2r+2}}{N} \sum_j \Delta F\left(\frac{j}{n}\right)\binom{r}{r+nx-j}^2 = \frac{n^{2r+2}}{N} \sum_{\nu=0}^{r} \Delta F\left(\frac{\nu}{n}+x\right)\binom{r}{r-\nu}^2 =$$

$$= \frac{n^{2r+1}}{N} f(x)4^r + \frac{1}{N} O(n^{2r}), \quad n \to \infty,$$

thereby finishing the proof of part (ii).

Looking at the case $\sigma^2(x)>0$ we see that the "optimal" choice $n=n(N)\sim \sim cN^{2/(2r+5)}$, $c>0$, $N\to\infty$ leads to the exact oıder of magnitude $N^{-4/(2r+5)}$ for the MSE of $\hat{f}_N^{(r)}$. Comparable results for the classical kernel estimator give the same rate of mean square convergence (e.g. [18]). Finally it should be pointed out that in particular the estimators (0.5) for the derivatives derived from (0.4) seem suitable rather than estimators obtained from $\hat{f}_N^{(0)}$ by differentiation with respect to x; for such estimators have complicated forms, if r is large. However in practice the computation of the coefficients of $p_{jn}(x)$ in (0.5) essentially requires only the evaluation of differences for a sequence of integers.

In this paper we have considered approximating operators and density estimators constructed by a lattice distribution. Motivated by a local central limit theorem another example is suggested by

$$p_{jn}(x) := \frac{1}{\sqrt{2\pi n}} e^{-(j-nx)^2/2n}; \quad j\in\mathbf{Z},\ x\in\mathbf{R},$$

which can be shown (see [16], [17]) to be a "good" approximation of a lattice distribution with mean nx and variance n (i.e. $\sigma^2(x)\equiv 1$). This approach leads to Favard operators for (0.1) [10], [17] for which the topics of this paper can be discussed in a similar way.

References

[1] M. Arató and A. Rényi, Probabilistic proof of a theorem of the approximation of continuous functions by means of generalized Bernstein polynomials, *Acta Math. Acad. Sci. Hungar.*, **8** (1957), 91—98.

[2] V. A. Baskakov, An example of a sequence of linear positive operators in the space of continuous functions, *Doklady Akad. Nauk. SSSR*, **113** (1957), 249—251.

[3] S. Bernstein, Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités, *Commun. Soc. Math. Kharkow*, (2) **13** (1912—13), 1—2.

[4] P. Butzer, Linear combinations of Bernstein polynomials, *Canad. J. Math.*, **5** (1953), 559—567.

[5] P. Butzer, On the extensions of Bernstein polynomials to the infinite interval, *Proc. Amer. Math. Soc.*, **5** (1954), 547—553.

[6] P. Butzer, Summability of generalized Bernstein polynomials. I, *Duke Math. J.*, **22** (1955), 617—622.

[7] Y. S. Chow and H. Teicher, *Probability theory, independence, interchangeability, martingales,* Springer-Verlag (New York—Heidelberg—Berlin, 1978).

[8] S. Dharmadhikari and K. Jogdeo, Bounds on moments of certain random variables, *Ann. Math. Statist.*, **40** (1969), 1506—1508.

[9] R. De Vore, *The approximation of continuous functions by positive linear operators,* Lecture notes in mathematics, No. 293, Springer-Verlag (Berlin—Heidelberg—New York, 1972).

[10] J. Favard, Sur les multiplicateurs d'interpolation, *J. Math. Pures Appl.*, **23** (1944), 219—247.

[11] F. Felbecker, Linearkombinationen von iterierten Bernsteinoperatoren, *Manuscripta Math.*, **29** (1979), 229—248.

[12] W. Feller, *An introduction to probability theory and its applications,* Vol. II, J. Wiley & Sons (New York—London—Sidney—Toronto, 1971).

[13] W. Gawronski, *Verallgemeinerte Bernsteinfunktionen und Schätzung einer Wahrscheinlichkeitsdichte,* Habilitationsschrift, Ulm, 1980.

[14] W. Gawronski, Strong laws for density estimators of Bernstein type, to appear.

[15] W. Gawronski and U. Stadtmüller, On density estimation by means of Poisson's distribution. *Scand. J. Statist.*, **7** (1980), 90—94.

[16] W. Gawronski and U. Stadtmüller, Smoothing histograms by means of lattice and continuous distributions, *Metrika*, **28** (1981), 155—164.

[17] W. Gawronski and U. Stadtmüller, Approximation of continuous functions by generalized Favard operators, *J. Approx. Theory*, **34** (1982), 384—396.

[18] T. Gasser and H.-G. Müller, Optimal convergence properties of kernel estimates of derivatives of a density function, in: *Smoothing techniques for curve estimation,* Lecture notes in mathematics, No. 757, Springer-Verlag (Berlin—Heidelberg—New York, 1979).

[19] S. Karlin and Z. Ziegler, Iteration of positive approximation operators, *J. Approx. Theory*, **3** (1970), 310—339.

[20] G. G. Lorentz, *Bernstein polynomials,* University of Toronto Press (Toronto, 1953).

[21] C. May, Saturation and inverse theorems for combinations of a class of exponential-type operators, *Canad. J. Math.*, **28** (1976), 1224—1250.

[22] M. W. Müller, *Approximationstheorie,* Akad. Verlagsgesellschaft (Wiesbaden, 1978).

[23] E. Parzen, On estimation of a probability density and mode, *Ann. Math. Statist.*, **33** (1962), 1065—1076.

[24] V. V. PETROV, *Sums of independent random variables*, Springer-Verlag (Berlin—Heidelberg—New York, 1975).

[25] P. RÉVÉSZ, *Gesetze der großen Zahlen*, Birkhäuser-Verlag (Basel—Stuttgart, 1968).

[26] R. S. SINGH, Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density, *Ann. Statist.*, **5** (1977), 394—399.

[27] U. STADTMÜLLER, Asymptotic distributions of smoothed histograms, *Metrika*, **30** (1983), 145—158.

[28] O. SZASZ, Generalization of S. Bernstein's polynomials to the infinite interval, *J. Research Nat. Bur. Stand.*, **45** (1950), 239—244.

[29] R. VITALE, A Bernstein polynomial approach to density estimation, in: *Statistical inference and related topics*, Academic Press (New York—San Francisco—London, 1975).

[30] H. WALK, *Approximation unbeschränkter Funktionen durch lineare positive Operatoren*, Habilitationsschrift, Stuttgart, 1970.

[31] H. WALK, Lokale Approximation unbeschränkter Funktionen und ihrer Ableitungen durch eine Klasse von Folgen linearer positiver Operatoren, *Mathematica*, **15** (1973), 129—142.

ABTEILUNG FÜR MATHEMATIK
UNIVERSITÄT ULM
OBERER ESELSBERG
7900 ULM (DONAU), FRG