On partial derivatives
of multivariate Bernstein polynomials[*]

A. Yu. Veretennikov[†], E. V. Veretennikova[‡]

October 16, 2016

Abstract

It is shown that Bernstein polynomials for a multivariate function converge to this function along with partial derivatives provided that the latter derivatives exist and are continuous. This result may be useful in some issues of stochastic calculus.

1 Introduction

Widely known is the proof of the polynomial Weierstrass theorem based on Bernstein polynomials and on the Law of Large Numbers for Bernoulli trials proposed in [3].

This method is applicable for the approximation of multivariate functions, too. In the literature it was noted that for a univariate function, Bernstein polynomials approximate this function along with its derivatives assuming that they exist, see for example [4, 5]. It turned out that this observation about an approximation along with derivatives holds true for multivariate functions as well; however, the authors were unable to find an exact reference and this was the reason for writing the present paper. In particular, this property is very useful in one of the proofs of

multi-dimensional Ito's formula, which formula is the main tool in stochastic analysis (cf. [6, Theorem 10.4]).

The latter proof starts with a verification of Ito's formula for the products; then it is extended by induction from linear functions to arbitrary polynomials and finally to all functions with two continuous and bounded derivatives; this approach essentially follows the lines of one of the proofs of the Weierstrass theorem (cf. [16, ch. XVI, §4]). It is essential that not only the function itself, but also its derivatives up to the second order were approximated on any compact. The latter requirement makes this version of the Weierstrass theorem a bit non-standard. At least, the majority of textbooks (cf. [16, 11]) where approximations by polynomials are discussed, usually focus on functions themselves and not on their derivatives.

In the univariate case this nuance is not important because we can approximate the n-th order derivative and then integrate it n times. But in the multivariate case this trick does not help because integrals may depend on the paths. In [6] the property of approximation of a function along with its derivatives is simply claimed as widely known. The authors found it interesting to inspect whether celebrated Bernstein polynomials, which gave rise to a notable branch of approximation theory $\big($cf. [9]$\big)$ admit this property. The paper [14] presents briefly the main result and the idea of approximation of partial derivatives in the case of $d = 2$. In this paper a full proof of this fact in the general case of $d \geq 2$ is provided. The case $d = 1$ is briefly presented in the next paragraph for the sake of completeness.

This paper consists of five sections. The first is Introduction. In the second some classical theorems and well-known generalizations in one-dimensional case are recalled. The third contains the main results about multivariate Bernstein polynomial derivatives convergence. The section 4 and 5 present the proofs of an auxiliary lemma and of the main results, correspondingly.

2 The case $d = 1$

For any function f on $[0, 1]$, approximating Bernstein polynomials are given by the formula
$$B_n(f; x) := \sum_{j=0}^{n} f\left(\frac{j}{n}\right) C_n^j x^j (1 - x)^{n-j}, \quad 0 \leq x \leq 1.$$

Theorem 1 (Bernstein) *If $f \in C([0, 1])$, then $B_n(f; x) \to f(x)$, $n \to \infty$ and this convergence is uniform $[0, 1]$.*

See [3], and some generalisations in [12]. The following similar result holds true for the derivatives.

2

Theorem 2 *If $f \in C^k([0,1])$, then $B_n^{(k)}(f;x) \to f^{(k)}(x)$, $n \to \infty$ and convergence is uniform on $[0,1]$.*

The proof of the Theorem 2 see, e.g., in [8, 7]; the result was established by I. N. Chlodovsky [4]. (The latter reference [4] gives only the title of his talk at the All-Unions Mathematical congress; the texts of most talks have not been published. According to [13, Chapter 1], since then the result of the Theorem 2 became known in the literature. However, the issue of correct references is a bit unclear.) There are various bounds of convergence rate under additional assumptions about smoothness or without them (cf. [15, 5, 10] et al.), but they are not the goal of this paper. The following identity for derivative will be useful in the sequel (cf., eg., [8, 7]):

$$B'_n(f;x) = \sum_{j=0}^{n-1} n\Delta_{1/n}f\left(\frac{j}{n}\right)C_{n-1}^j x^j (1-x)^{n-1-j}, \tag{1}$$

where

$$\Delta_z f\left(\frac{j}{n}\right) := f\left(\frac{j}{n} + z\right) - f\left(\frac{j}{n}\right), \quad z \in \mathbb{R}. \tag{2}$$

Later for a multi-dimensional case we will use a more detailed notation Δ_{z,x^i}, which emphasises that the increment corresponds to the variable x^i; for a fixed $z = 1/n$ we will use a short notation $\Delta_{(x^i)}$. By induction the representation of the derivative of order k follows:

$$B_n^{(k)}(f;x) = \frac{n!}{(n-k)!} \sum_{j=0}^{n-k} \Delta_{1/n}^k f\left(\frac{j}{n}\right) C_{n-k}^j x^j (1-x)^{n-k-j}, \tag{3}$$

where Δ_z^k is defined also by induction as the operator Δ_z applied k times. For example for $k = 2$ we have,

$$\Delta_z^2 f(x) = \Delta_z\big(\Delta_z f(x)\big) = f(x + 2z) - 2f(x + z) + f(x).$$

Note that the latter expression is one of the versions of a non-normalized finite difference Laplacian for the one-dimensional case. Further, since $n^k \Delta_{1/n}^k f(x) \rightrightarrows f^{(k)}(x) \frac{n!}{(n-k)!}/n^k \to 1$, $n \to \infty$, under any fixed k, then by virtue of (3) and due to the main calculus in the proof of the Theorem 1 based on the law of large numbers in the Bernoulli trials scheme with probability of success x (we recall that $x \in [0,1]$), the proof of convergence for the derivatives follows immediately, of course, under the assumption $f \in C^k(\mathbb{R})$. It should be noted that all convergences are uniform

3

on $[0,1]$. We do not show here the details of this well-known calculation $\big($cf. [8, Theorem 1.8.1] in the case $k = 1\big)$ as for the main result of this paper in the multi-dimensional case they all will be given in the next sections. Yet we want to point out the development of the key Bernstein's idea, which suggests to reformulate (1) and (3), respectively, as

$$B_n'(f; x) = \mathbb{E} n \Delta_{1/n} f\big(n^{-1}\xi_{n-1}(x)\big)$$

and

$$B_n^{(k)}(f; x) = \mathbb{E} \frac{n!}{(n-k)!} \Delta_{1/n}^k f\big(n^{-1}\xi_{n-k}(x)\big),$$

where $\xi_n(x)$ denotes a random variable with Binomial distribution $\mathrm{Bin}(n, x)$. These representations clearly explain why in the case $d = 1$ the left hand sides of (1) and (3) tend by the law of large numbers to their limits $f'(x)$ and $f^{(k)}(x)$, respectively, under the assumptions of the Theorem 2.

3   The case $d \geq 2$ — Main Result

Now let us consider $d \geq 2$, $x \in \mathbb{R}^d$, $x = (x_1, \ldots, x_d)$. There are at least two ways to define Bernstein polynomials in the multi-dimensional case (actually, there are many ways and we will describe them later): either on a simplex $\mathbb{S}^d := \big((x_1, x_2, \ldots, x_d) : 0 \leq x_1, x_2, \ldots, x_d, \ \|x\| \leq 1\big)$ by the formula

$$B_n(f; x) := \sum_{\substack{0 \leq j_1 + j_2 + \cdots + j_d \leq n, \\ j_1, j_2, \ldots, j_d \geq 0}} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right) C_n^j x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} \big(1 - \|x\|\big)^{n - |j|}, \quad (4)$$

where $j = (j_1, j_2, \ldots, j_d)$, and where a vector x norm, and the "modulus" of a multi-index j, and the polynomial coefficient ("n choose j") C_n^j are defined as $\|x\| = |x_1| + |x_2| + \cdots + |x_d|$, $|j| = j_1 + j_2 + \cdots + j_d$, and

$$C_n^j \equiv C_n^{j_1, \ldots, j_d} = \frac{n!}{j_1! j_2! \ldots j_d! (n - |j|)!};$$

or on a "d-dimensional square" (cube, etc.) by a formula similar to but different from (4),

$$\widetilde{B}_n(f; x) := \sum_{j_1, j_2, \ldots, j_d = 0}^{n} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right) C_n^{j_1} C_n^{j_2} \cdots C_n^{j_d}$$

$$\times x_1^{j_1} (1 - x_1)^{n - j_1} x_2^{j_2} (1 - x_2)^{n - j_2} \cdots x_d^{j_d} (1 - x_d)^{n - j_d}. \quad (5)$$

4

Note that the degrees of the polynomials B_n and $\tilde{B}_n$ are different. The formulae (4) and (5) allow probabilistic representation

$$B_n(f;x) := \mathbb{E}f\big(n^{-1}\eta_n(x)\big) \tag{6}$$

and

$$\widetilde{B}_n(f;x) := \mathbb{E}f\big(n^{-1}\xi_n(x)\big), \tag{7}$$

where the distribution of the random vector $\eta_n(x)$ of dimension d is the projection onto the first d coordinates of the *polynomial (multinomial) distribution* of dimension $d+1$ with parameters n (the number of trials) and the vector of "success probabilities" $(x_1, \ldots, x_d, x_{d+1})$, which satisfy conditions $x_i \geq 0$ $(i = 1, \ldots, d+1)$ and $\sum_{i=1}^{d+1} x_i = 1$, while the random vector $\xi_n(x) = (\xi_n^1, \ldots, \xi_n^d)$ consists of independent components distributed binomially $\mathrm{Bin}(n, x_i)$ each.

Due to the law of large numbers for Bernoulli trials for each coordinate we have,

$$\frac{1}{n}\xi_n(x) \to x, \quad n \to \infty, \tag{8}$$

both in probability and almost surely. For the sequence (η_n) the law of large numbers also holds true:

$$\frac{1}{n}\eta_n(x) \to x, \quad n \to \infty, \tag{9}$$

in probability and almost surely. The easiest way to show this is to use the law of large numbers for each coordinate, which follows directly from the one-dimensional version of this theorem for the binomial distribution.

It apparently does not help too much to analyse partial derivatives where certain series still need to be treated. However, after the differentiation of these series we will get expressions, which admit representations via expectations of finite differences for the function f with random vector arguments distributed either according to a polynomial law – or, more precisely, its projection – or as a direct product of binomial distributions, which will eventually lead to the desired result.

Remark 1 *Other variants are possible, such as a combination of a "cube" and of a simplex for different variables. For example, in case of $d = d_1 + d_2$ and $d_1 \geq 2$ one more version of multivariate Bernstein polynomials has a form,*

$$\widehat{B}_n(f;x) := \sum_{\substack{0 \le j_1+j_2+\cdots+j_{d_1} \le n, \\ j_1,j_2,\ldots,j_{d_1} \ge 0}} f\left(\frac{j_1}{n}, \ldots, \frac{j_{d_1}}{n}, \frac{j_{d_1+1}}{n}, \ldots, \frac{j_{d_1+d_2}}{n}\right)$$

$$\times C_n^{j_1,\ldots,j_{d_1}} x_1^{j_1} x_2^{j_2} \cdots x_{d_1}^{j_{d_1}} \left(1 - \sum_{i=1}^{d_1} x_i\right)^{n-j_1-\cdots-j_{d_1}}$$

$$\times C_n^{j_{d_1+1}} \cdots C_n^{j_{d_1+d_2}} x_{d_1+1}^{j_{d_1+1}} (1-x_1)^{n-j_{d_1+1}} \cdots x_{d_1+d_2}^{j_{d_1+d_2}} (1-x_{d_1+d_2})^{n-j_{d_1+d_2}},$$

where $x = (x_1, \ldots, x_{d_1}, x_{d_1+1}, \ldots, x_{d_1+d_2})$, all x_i are non-negative, $x_1 + \cdots + x_{d_1} \le 1$, and $0 \le x_{d_1+1} \le 1, \ldots, 0 \le x_{d_1+d_2} \le 1$.

It is quite likely that for such representations analogous results about convergence of polynomials and their derivatives may be established.

So, a multivariate analogue of the Theorem 1 for the polynomials B_n and $\tilde{B}_n$ may be formulated as follows.

Theorem 3 *If $f \in C(\mathbb{R}^d)$, then*

$$\tilde{B}_n(f;x) \to f(x), \quad x \in \mathbb{K}^d,$$
$$B_n(f;x) \to f(x), \quad x \in \mathbb{S}^d,$$

as $n \to \infty$. All convergences are uniform on $\mathbb{K}^d$ and $\mathbb{S}^d$, respectively.

Note that any continuous function on $\mathbb{S}^d$ or on $\mathbb{K}^d$ can be extended to a continuous function on $\mathbb{R}^d$ (of course, not uniquely). Convergence of both versions towards a continuous function f on the simplex or on the d-dimensional square/cube follows from the (multivariate) law of large numbers: in the second case it is applied to a sequence of independent and equally distributed random vectors with independent components, while in the first case – with dependent components. The proof can be easily found in may papers and textbooks (cf. e.g., [5]) and we omit it.

To state the main result, let us introduce the following notations: :

1. m — a natural number;

2. $k = (k_1, k_2, \ldots, k_d)$ — a multi-index;

3. C^k *with a multi-index k* k denotes the class of functions that possess a mixed partial derivative of order $k = (k_1, k_2, \ldots, k_d)$, which is continuous.

Recall that $C^m(\mathbb{R}^d)$ $\left(C_b^m(\mathbb{R}^d)\right)$ with $m = 1, 2, \ldots$ is a class of functions on $\mathbb{R}^d$ with *all* well-defined mixed derivatives of order m, which are continuous (respectively, continuous and bounded). The notation $f^{(m)}$ stands for the set of all mixed derivatives of the function f of order m.

Theorem 4 1. *If $m > 0$, $f \in C^m(\mathbb{R}^d)$, then*

$$\widetilde{B}_n^{(m)}(f; x) \to f^{(m)}(x), \quad x \in \mathbb{K}^d,$$
$$B_n^{(m)}(f; x) \to f^{(m)}(x), \quad x \in \mathbb{S}^d,$$

as $n \to \infty$, and all convergences are uniform on $\mathbb{K}^d$ and $\mathbb{S}^d$, respectively.
  2. If $k = (k_1, k_2, \ldots, k_d)$ is a multi-index and $f \in C^k(\mathbb{R}^d)$, then

$$\widetilde{B}_n^{(k)}(f; x) \to f^{(k)}(x), \quad x \in \mathbb{K}^d,$$
$$B_n^{(k)}(f; x) \to f^{(k)}(x), \quad x \in \mathbb{S}^d,$$

as $n \to \infty$, and all convergences are uniform on $\mathbb{K}^d$ and $\mathbb{S}^d$, respectively.

It is worth mentioning also the papers [1] and [2] devoted to asymptotic expansions and Taylor's expansions for Bernstein polynomials of two and many variables. While being conceptually close, the results of the present paper do not follow directly from these papers, and we do not aim to get neither asymptotic nor Taylor's expansions here.

4 Auxiliary result

The following Lemma will be used in the proof of the main Theorem when a uniform convergence of finite differences just under the condition of existence and continuity of the limit expression is needed. While elementary, this Lemma is required for the correctness of references and for a completeness of our presentation.

Let $(z_1, \ldots, z_d) \in \mathbb{R}^d$, and $(x_1, \ldots, x_d) \in \mathbb{R}^d$. Let

$$\Delta_{1/n, x_i} f(x) := f(x_1, \ldots, x_{i-1}, x_i + \frac{1}{n}, x_{i+1}, \ldots, x_d)$$
$$- f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d),$$
$$\partial^k f(x_1, \ldots, x_d) := \frac{\partial^{|k|}}{\partial_{x_1}^{k_1} \ldots \partial_{x_d}^{k_d}} f(x_1, \ldots, x_d).$$

Lemma 1 *For $f \in C^k(\mathbb{R}^d)$ with $k = (k_1, \ldots, k_d)$ and $x = (x_1, \ldots, x_d)$ the following equality is valid*

$$\Delta_{z_1,x_1}^{k_1} \ldots \Delta_{z_d,x_d}^{k_d} f(x) = \int_{x_1}^{x_1+z_1} d\xi_1^1 \int_{\xi_1^1}^{\xi_1^1+z_1} d\xi_2^1 \ldots \int_{\xi_{k_1-1}^1}^{\xi_{k_1-1}^1+z_1} d\xi_{k_1}^1$$

$$\ldots \int_{x_d}^{x_d+z_d} d\xi_1^d \int_{\xi_1^d}^{\xi_1^d+z_d} d\xi_2^d \ldots \int_{\xi_{k_d-1}^d}^{\xi_{k_d-1}^d+z_d} d\xi_{k_d}^d \partial^k f(\xi_{k_1}^1, \ldots, \xi_{k_d}^d).$$

For example, for $k_1 = 0$, integration over ξ_1^1, ξ_2^1, etc. is just not performed, and $f(\xi_{k_1}^1, \xi_{k_2}^2, \ldots, \xi_{k_d}^d)$ is treated as $f(x_1, \xi_{k_2}^2, \ldots, \xi_{k_d}^d)$. In particular, if all $k_i = 0$ then the equality turns into the identity $f(x) = f(x)$.

The equality for two variables $x = (x_1, x_2)$ and $k_1 = k_2 = 1$ takes the form (for the sake of simplicity we omit the lower indices in ξ_1^1, ξ_1^2):

$$\Delta_{z_1,x_1} \Delta_{z_2,x_2} f(x) = \int_{x_1}^{x_1+z_1} d\xi^1 \int_{x_2}^{x_2+z_2} d\xi^2 \frac{\partial^2}{\partial x_1 \partial x_2} f(\xi^1, \xi^2).$$

The proof of the Lemma follows straightforward from the (one-dimensional) first theorem of the calculus (also known as Newton-Leibniz formula) by induction.

5 Proof of the Theorem 4

0. It suffices to prove only the second part of the Theorem 4 because the first part follows immediately due to the identity

$$C^m(\mathbb{R}^d) = \bigcap_{k:\ |k| \leq m} C^k(\mathbb{R}^d).$$

1. Analogues of the formulae (1) and (2) in the multivariate case for $\widetilde{B}_n$ for the

multi-index $k = (k_1, \ldots, k_d)$ could be written as

$$
\frac{\partial^{|k|}}{\partial x^k} \widetilde{B}_n(f; x) \equiv \frac{\partial^{|k|}}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}} \widetilde{B}_n(f; x)
$$

$$
= \left(\frac{n \ldots (n - k_1 + 1) \times n \ldots (n - k_2 + 1) \times \cdots \times n \ldots (n - k_d + 1)}{n^{|k|}} \right)
$$

$$
\times \sum_{\substack{0 \leq j_i \leq n - k_i, \\ i = 1, 2, \ldots, d}} n^{|k|} \Delta^k f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n} \right) C_{n-k_1}^{j_1} C_{n-k_2}^{j_2} \cdots C_{n-k_d}^{j_d}
$$

$$
\times x_1^{j_1}(1 - x_1)^{n-k_1-j_1} x_2^{j_2}(1 - x_2)^{n-k_2-j_2} \cdots x_d^{j_d}(1 - x_d)^{n-k_d-j_d}, \tag{10}
$$

where

$$
\frac{\partial^{|k|}}{\partial x^k} := \frac{\partial^{|k|}}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}}, \quad \Delta^k := \prod_{i=1}^{d} \Delta_{(x_i)}^{k_i},
$$

and where $\Delta_{(x_i)}$ means $\Delta_{(x_i)} \equiv \Delta_{1/n, x_i}$.

The usage of the notation Δ^k with a multi-index k is correct because all operators $\Delta_{(x_i)}$ commute. Indeed, let us denote (assume $i < j$ for the sake of definiteness)

$$
x^{i,j} = \left(x^1, \ldots, x^{i-1}, x^i + 1/n, x^{i+1}, \ldots, x^{j-1}, x^j + 1/n, \ldots, x^d \right),
$$

$$
x^i = \left(x^1, \ldots, x^{i-1}, x^i + 1/n, x^{i+1}, \ldots, x^d \right),
$$

$$
x^j = \left(x^1, \ldots, x^{j-1}, x^j + 1/n, \ldots, x^d \right).
$$

For $i \neq j$ we get elementary identities:

$$
\begin{aligned}
\Delta_{(x_i)}\left(\Delta_{(x_j)} f(x) \right) &= \Delta_{(x_i)}\left(f(x^j) - f(x) \right) \\
&= f(x^{i,j}) - f(x^i) - \left(f(x^j) - f(x) \right) \\
&= f(x^{i,j}) - f(x^i) - f(x^j) + f(x)
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
\Delta_{(x_j)}\left(\Delta_{(x_i)} f(x) \right) &= \Delta_{(x_j)}\left(f(x^i) - f(x) \right) \\
&= f(x^{i,j}) - f(x^j) - \left(f(x^i) - f(x) \right) \\
&= f(x^{i,j}) - f(x^i) - f(x^j) + f(x).
\end{aligned}
$$

Analogously to the representation (6), the equality (10) admits the following probabilistic meaning:

$$\frac{\partial^{|k|}}{\partial x^k}\widetilde{B}_n(f;x) = \left(\frac{n\ldots(n-k_1+1)\times n\ldots(n-k_2+1)\times\cdots\times n\ldots(n-k_d+1)}{n^{|k|}}\right)$$

$$\times \mathbb{E}n^{|k|}\Delta^k f(n^{-1}\xi_{n-k}(x)); \tag{11}$$

the definition of the random vector $\xi_n(x) = (\xi_n^1,\ldots,\xi_n^d)$ see in § 3).

2. Let us prove the equality (10) by induction. For $k = (0,\cdots,0)$ (a multi-index of order d) the desired equality (3) is equivalent to the definition of the polynomial $\tilde{B}_n(f;x)$; this will serve as the *basis of induction*. It should be noted that for several variables (d in our case) induction can be carried in turn for the first variable up to k_1, then for the second one up to k_2, and so on.

In other words, "double" induction (it can be also named multivariate) over each of the variables x_i, and then over the indices $i = 1,\ldots,d$ may be applied in our situation. For the induction step, it actually suffices to verify that the formula (3) remains valid when any component of the multi-index $k = (k_1,\cdots,k_d)$ increases by one. For the sake of definiteness let us check the step $k_i \mapsto k_i + 1$ for $i = 1$. Recall that

$$\widetilde{B}_n(f;x) := \sum_{j_1,j_2,\ldots,j_d=0}^{n} f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right) C_n^{j_1}C_n^{j_2}\cdots C_n^{j_d}$$

$$\times x_1^{j_1}(1-x_1)^{n-j_1}x_2^{j_2}(1-x_2)^{n-j_2}\cdots x_d^{j_d}(1-x_d)^{n-j_d}.$$

Because of a certain clumsiness of the formulae and for the sake of clarity we shall start with $k_i = 0$ and then proceed to the general case. (The same approach will be used for the polynomials B_n). We get,

$$\partial_{x_1} \widetilde{B}_n(f; x) = \partial_{x_1} \sum_{j_1, j_2, \ldots, j_d = 0}^{n} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right) C_n^{j_1} C_n^{j_2} \cdots C_n^{j_d}$$

$$\times x_1^{j_1} (1-x_1)^{n-j_1} x_2^{j_2} (1-x_2)^{n-j_2} \cdots x_d^{j_d} (1-x_d)^{n-j_d}$$

$$= \sum_{\substack{0 < j_1 \leq n, \\ 0 \leq j_2, \ldots, j_d \leq n}} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right) C_n^{j_1} C_n^{j_2} \cdots C_n^{j_d}$$

$$\times j_1 x_1^{j_1 - 1} (1-x_1)^{n-j_1} x_2^{j_2} (1-x_2)^{n-j_2} \cdots x_d^{j_d} (1-x_d)^{n-j_d}$$

$$- \sum_{\substack{0 \leq j_1 < n, \\ 0 \leq j_2, \ldots, j_d \leq n}} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right) C_n^{j_1} C_n^{j_2} \cdots C_n^{j_d}$$

$$\times x_1^{j_1} (n-j_1)(1-x_1)^{n-j_1-1} x_2^{j_2} (1-x_2)^{n-j_2} \cdots x_d^{j_d} (1-x_d)^{n-j_d}$$

$$= \sum_{\substack{0 \leq j_1 < n, \\ 0 \leq j_2, \ldots, j_d \leq n}} f\left(\frac{j_1 + 1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right) C_n^{j_1 + 1} C_n^{j_2} \cdots C_n^{j_d}$$

$$\times (j_1 + 1) x_1^{j_1} (1-x_1)^{n-(j_1+1)} x_2^{j_2} (1-x_2)^{n-j_2} \cdots x_d^{j_d} (1-x_d)^{n-j_d}$$

$$- \sum_{\substack{0 \leq j_1 < n, \\ 0 \leq j_2, \ldots, j_d \leq n}} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right) C_n^{j_1} C_n^{j_2} \cdots C_n^{j_d}$$

$$\times x_1^{j_1} (n-j_1)(1-x_1)^{n-j_1-1} x_2^{j_2} (1-x_2)^{n-j_2} \cdots x_d^{j_d} (1-x_d)^{n-j_d}$$

$$= \sum_{\substack{0 \leq j_1 < n, \\ 0 \leq j_2, \ldots, j_d \leq n}} n \Delta_{(x_1)} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right) C_{n-1}^{j_1} C_n^{j_2} \cdots C_n^{j_d}$$

$$\times x_1^{j_1} (1-x_1)^{n-1-j_1} x_2^{j_2} (1-x_2)^{n-j_2} \cdots x_d^{j_d} (1-x_d)^{n-j_d},$$

due to the identities

$$(j_1 + 1) C_n^{j_1 + 1} = \frac{(j_1 + 1) n!}{(j_1 + 1)! \left(n - (j_1 + 1)\right)!} = \frac{n!}{j_1! (n - 1 - j_1)!} = n C_{n-1}^{j_1},$$

$$(n - j_1) C_n^{j_1} = \frac{(n - j_1) n!}{j_1! (n - j_1)!} = \frac{n!}{j_1! (n - 1 - j_1)!} = n C_{n-1}^{j_1}.$$

Now, assuming that the formula (3) holds true for some $k = (k_1, \ldots, k_d)$, let us differentiate it once more in variable x_1 so as to get a similar formula for the

multi-index $(k_1 + 1, k_2, \ldots, k_d)$. For the sake of brevity let us denote

$$\alpha_{k_1,k_2,\ldots,k_d}^{j_1,j_2,\ldots,j_d} := C_{n-k_1}^{j_1} \cdots C_{n-k_d}^{j_d} x_1^{j_1} (1-x_1)^{n-k_1-j_1} \cdots x_d^{j_d} (1-x_d)^{n-k_d-j_d},$$

$$\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d} := C_{n-k_2}^{j_2} \cdots C_{n-k_d}^{j_d} x_2^{j_2} (1-x_2)^{n-k_2-j_2} \cdots x_d^{j_d} (1-x_d)^{n-k_d-j_d}.$$

Note that since we are interested in the statement of the Theorem for $n \to \infty$, we can assume that all $k_i \le n$ $\big($although, for $k_i > n$ we have $\partial_{x_i}^{k_i} \tilde{B}_n(x) = 0$, and the right side of (3) also equals zero by definition of the number of combinations with negative $n - k_i\big)$.

Derivative of a constant being equal to zero, we have,

$$\frac{\partial}{\partial x_1}\left(\sum_{\substack{0\le j_i\le n-k_i,\\ i=1,\ldots,d}}\Delta^k f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)\alpha_{k_1,\ldots,k_d}^{j_1,\ldots,j_d}\right)$$

$$=\frac{\partial}{\partial x_1}\left(\sum_{\substack{0\le j_i\le n-k_i,\\ i=1,\ldots,d}}\Delta^k f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)C_{n-k_1}^{j_1}\,x_1^{j_1}(1-x_1)^{n-k_1-j_1}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}\right)$$

$$=\frac{\partial}{\partial x_1}\sum_{j_1=0}^{n-k_1}\sum_{\substack{0\le j_i\le n-k_i,\\ i=2,\ldots,d}}\Delta^k f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)C_{n-k_1}^{j_1}\,x_1^{j_1}(1-x_1)^{n-k_1-j_1}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}$$

$$=\sum_{j_1=1}^{n-k_1}\sum_{\substack{0\le j_i\le n-k_i,\\ i=2,\ldots,d}}\Delta^k f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)C_{n-k_1}^{j_1}\,j_1 x_1^{j_1-1}(1-x_1)^{n-k_1-j_1}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}$$

$$-\sum_{j_1=0}^{n-k_1-1}\sum_{\substack{0\le j_i\le n-k_i,\\ i=2,\ldots,d}}\Delta^k f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)$$
$$\times C_{n-k_1}^{j_1}\,x_1^{j_1}(n-k_1-j_1)(1-x_1)^{n-k_1-j_1-1}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}$$

$$=\sum_{\substack{0\le j_i\le n-k_i,\\ i=2,\ldots,d}}\sum_{j_1'=0}^{n-k_1-1}\Delta^k f\left(\frac{j_1'+1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)$$
$$\times C_{n-k_1}^{j_1'+1}\,(j_1'+1)x_1^{j_1'}(1-x_1)^{n-k_1-j_1'-1}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}$$

$$-\sum_{\substack{0\le j_i\le n-k_i,\\ i=2,\ldots,d}}\sum_{j_1=0}^{n-k_1-1}\Delta^k f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)$$
$$\times C_{n-k_1}^{j_1}\,x_1^{j_1}(n-k_1-j_1)(1-x_1)^{n-k_1-j_1-1}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}$$

$$=\sum_{\substack{0\le j_i\le n-k_i,\\ i=2,\ldots,d}}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}\left(\sum_{j_1=0}^{n-k_1-1}\Delta^k f\left(\frac{j_1+1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)\right.$$
$$\times C_{n-k_1}^{j_1+1}(j_1+1)x_1^{j_1}(1-x_1)^{n-k_1-j_1-1}-\Delta^k f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)$$
$$\left.\times C_{n-k_1}^{j_1}\,x_1^{j_1}(n-k_1-j_1)(1-x_1)^{n-k_1-j_1-1}\right)$$

$$=\sum_{\substack{0\le j_i\le n-k_i,\\ i=2,\ldots,d}}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}\left(\sum_{j_1=0}^{n-k_1-1}(n-k_1)C_{n-(k_1+1)}^{j_1}x_1^{j_1}(1-x_1)^{n-(k_1+1)-j_1}\right.$$
$$\left.\times\left(\Delta^k f\left(\frac{j_1+1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)-\Delta^k f\left(\frac{j_1}{n},\frac{j_2}{n},\ldots,\frac{j_d}{n}\right)\right)\right)$$

$$\equiv\sum_{\substack{0\le j_i\le n-k_i,\\ i=2,\ldots,d}}\alpha_{k_2,\ldots,k_d}^{j_2,\ldots,j_d}\sum_{j_1=0}^{n-k_1-1}(n-k_1)$$

as required. We have used the identities

$$C_{n-k_1}^{j_1+1}(j_1+1) = \frac{(n-k_1)!(j_1+1)}{(j_1+1)!(n-k_1-j_1-1)!}$$
$$= \frac{(n-k_1)(n-k_1-1)!}{j_1!(n-k_1-j_1-1)!} = (n-k_1)\,C_{n-(k_1+1)}^{j_1},$$
$$C_{n-k_1}^{j_1}(n-k_1-j_1) = \frac{(n-k_1-j_1)(n-k_1)!}{j_1!(n-k_1-j_1)!} = (n-k_1)\,C_{n-(k_1+1)}^{j_1}.$$

Hence, it follows by induction that the formula (3) holds true.

3. As for $\widetilde{B}_n$, it remains to note that for $f \in C^k(\mathbb{R}^d)$ due to the Lemma 1, appropriately normalized finite differences are uniformly close to the corresponding partial derivatives, that is,

$$n^{|k|}\Delta^k f(x) \rightrightarrows \frac{\partial^{|k|} f}{\partial x^k}(x) \quad x \in \mathbb{K}^d, \quad n \to \infty.$$

So, the statement of the Theorem for the polynomials $\widetilde{B}_n$ follows from (11) and from the law of large numbers (8).

4. The analogue of the formula (3) for the polynomials B_n has a form,

$$B_n^{(k)}(f;x) \equiv \frac{\partial^{|k|}}{\partial x^k} B_n(f;x) \equiv \frac{\partial^{|k|}}{\partial x_1^{k_1} \partial x_2^{k_2} \dots \partial x_d^{k_d}} B_n(f;x)$$
$$= \sum_{0 \le j_1+j_2+\dots \le n-|k|} \frac{n!}{(n-|k|)!} \Delta_{(x_1)}^{k_1} \Delta_{(x_2)}^{k_2} \dots \Delta_{(x_d)}^{k_d}$$
$$\times f\left(\frac{j_1}{n}, \frac{j_2}{n}, \dots, \frac{j_d}{n}\right) C_{n-|k|}^{j_1,j_2,\dots,j_d} x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} (1-\|x\|)^{n-|k|-|j|}. \qquad (12)$$

For the seequel, let us recall the following notation,

$$\Delta_{(x_1)}^{k_1} \Delta_{(x_2)}^{k_2} \dots \Delta_{(x_d)}^{k_d} = \Delta^k.$$

The formula (12) also has a simple probabilistic meaning. Similarly to the representation (7) we write,

$$B_n^{(k)}(f;x) = \mathbb{E}\frac{n!}{(n-|k|)!} \Delta^k f\left(n^{-1}\eta_{n-k}(x)\right); \qquad (13)$$

for the definition of a random vector $\eta_n(x)$ see §3.

14

5. For the proof by induction of the relations (12) and (13), let us note that the basis of induction ($k = 0$) coincides with the definition of the polynomial $B_n(f;x)$; as for the inductive step, for the sake of clarity of the following bulky computations let us first differentiate once the function $B_n(f;x)$. we have,

$$\partial_{x_1} B_n(f;x) := \sum_{\substack{0 \le j_1+j_2+\cdots+j_d \le n, \\ j_1,j_2,\ldots j_d \ge 0}} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_n^j \partial_{x_1} x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d}(1-x_1-x_2-\cdots-x_d)^{n-|j|}$$

$$= \sum_{\substack{0 \le j_1+j_2+\cdots+j_d \le n, \\ j_1>0,\, j_2,\ldots,j_d \ge 0}} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_n^j j_1 x_1^{j_1-1} x_2^{j_2} \cdots x_d^{j_d}(1-x_1-x_2-\cdots-x_d)^{n-|j|}$$

$$- \sum_{\substack{0 \le j_1+j_2+\cdots+j_d < n, \\ j_1,j_2,\ldots,j_d \ge 0}} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_n^j \left(n-|j|\right) x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d}(1-x_1-x_2-\cdots-x_d)^{n-|j|-1}$$

$$= \sum_{\substack{0 \le j_1+j_2+\cdots+j_d \le n-1, \\ j_1,j_2,\ldots,j_d \ge 0}} f\left(\frac{j_1+1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_n^{j_1+1,\ldots,j_d}(j_1+1) x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d}(1-x_1-x_2-\cdots-x_d)^{n-|j|}$$

$$- \sum_{\substack{0 \le j_1+j_2+\cdots+j_d < n, \\ j_1,j_2,\ldots,j_d \ge 0}} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_n^{j_1,\ldots,j_d}\left(n-|j|\right) x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d}(1-x_1-x_2-\cdots-x_d)^{n-1-|j|}$$

$$= \sum_{\substack{0 \le j_1+j_2+\cdots+j_d < n, \\ j_1,j_2,\ldots,j_d \ge 0}} n\Delta_{(x_1)} f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_{n-1}^{j_1,\ldots,j_d} x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d}(1-x_1-x_2-\cdots-x_d)^{n-1-|j|},$$

due to the identities

$$C_n^{j_1+1,j_2,\ldots,j_d}(j_1+1) = \frac{n!(j_1+1)}{(j_1+1)!j_2!\ldots} = n\frac{(n-1)!}{j_1!j_2!\cdots j_d!} = nC_{n-1}^{j_1,j_2,\ldots,j_d},$$

$$C_n^{j_1,j_2,\ldots,j_d}(n-j_1-\cdots-j_d) = \frac{n!(n-j_1-\cdots-j_d)}{j_1!j_2!\cdots(n-|j|)!} = nC_{n-1}^{j_1,j_2,\ldots,j_d}.$$

6. The "full" induction step: under the assumption that the formula (12) holds true for some multi-index $k = (k_1, \ldots, k_d)$, let us differentiate this expression again,

say, with respect to x_1. We get,

$$\partial_{x_1} B_n^{(k)}(f;x) = \partial_{x_1} \sum_{0 \le j_1+j_2+\cdots+j_d \le n-|k|} \frac{n!}{(n-|k|)!} \Delta^k f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_{n-|k|}^{j_1,j_2,\ldots,j_d} x_1^{j_1} x_2^{j_2} \cdots \left(1-\|x\|\right)^{n-|k|-|j|}$$

$$= \sum_{\substack{0 \le j_1+j_2+\cdots+j_d \le n, \\ j_1 > 0, j_2, \ldots, j_d \ge 0}} \frac{n!}{(n-|k|)!} \Delta^k f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_{n-|k|}^{j} j_1 x_1^{j_1-1} x_2^{j_2} \cdots x_d^{j_d} (1-x_1-x_2-\cdots-x_d)^{n-|k|-|j|}$$

$$- \sum_{\substack{0 \le j_1+j_2+\cdots+j_d < n, \\ j_1, j_2, \ldots, j_d \ge 0}} \frac{n!}{(n-|k|)!} \Delta^k f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_{n-|k|}^{j} \left(n-|k|-|j|\right) x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} (1-x_1-x_2-\cdots-x_d)^{n-|k|-|j|-1}$$

$$= \sum_{\substack{0 \le j_1+j_2+\cdots+j_d \le n-1, \\ j_1, j_2, \ldots j_d \ge 0}} \frac{n!}{(n-|k|)!} \Delta^k f\left(\frac{j_1+1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_{n-|k|}^{j_1+1,\ldots,j_d} (j_1+1) x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} (1-x_1-x_2-\cdots-x_d)^{n-|k|-|j|}$$

$$- \sum_{\substack{0 \le j_1+j_2+\cdots+j_d < n, \\ j_1, j_2, \ldots j_d \ge 0}} \frac{n!}{(n-|k|)!} \Delta^k f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_{n-|k|}^{j_1,\ldots,j_d} \left(n-|k|-|j|\right) x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} (1-x_1-x_2-\cdots-x_d)^{n-|k|-1-|j|}$$

$$= \sum_{\substack{0 \le j_1+j_2+\cdots+j_d < n, \\ j_1, j_2, \ldots j_d \ge 0}} \left(n-|k|\right) \frac{n!}{(n-|k|)!} \Delta_{(x_1)} \Delta^k f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_{n-|k|-1}^{j_1,\ldots,j_d} x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} (1-x_1-x_2-\cdots-x_d)^{n-|k|-1-|j|}$$

$$= \sum_{\substack{0 \le j_1+j_2+\cdots+j_d < n, \\ j_1, j_2, \ldots j_d \ge 0}} \frac{n!}{(n-|k|-1)!} \Delta_{(x_1)} \Delta^k f\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_d}{n}\right)$$

$$\times C_{n-|k|-1}^{j_1,\ldots,j_d} x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} (1-x_1-x_2-\cdots-x_d)^{n-|k|-1-|j|},$$

by virtue of the identities

$$C_{n-|k|}^{j_1+1,j_2,\ldots,j_d}(j_1+1) = \frac{(n-|k|)!(j_1+1)}{(j_1+1)!j_2!\cdots(n-|k|-|j|-1)!}$$
$$= (n-|k|)\frac{(n-|k|-1)!}{j_1!j_2!\cdots(n-|k|-1-|j|)!}$$
$$= (n-|k|)C_{n-|k|-1}^{j_1,j_2,\ldots,j_d},$$
$$C_{n-|k|}^{j_1,j_2,\ldots,j_d}\big(n-|k|-j_1-\cdots-j_d\big) = \frac{(n-|k|)!(n-|k|-|j|)}{j_1!j_2!\cdots,j_d!(n-|k|-|j|)!}$$
$$= (n-|k|)C_{n-|k|-1}^{j_1,j_2,\ldots,j_d}.$$

Hence, by induction the formula (12) holds true along with (13).

7. Again having in mind how Bernstein's method for the Theorem 1 was applied in the proof of the Theorem 4 for the polynomials B_n, and similarly to what was done earlier for $\widetilde{B}_n$, let us recall that for $f \in C^k(\mathbb{R}^d)$ the normalized finite differences are uniformly close to the corresponding partial derivatives, namely,

$$\frac{n!}{(n-|k|)!}\,\Delta^k f(x) \rightrightarrows \frac{\partial^{|k|}f}{\partial x^k}(x), \quad \text{for} \quad x \in \mathbb{S}^d, \quad n \to \infty$$

(cf. the Lemma 1). Hence statement of the Theorem for the polynomials B_n follows from the law of large numbers (9) and from the representation (13).

The Theorem 4 is proved

Acknowledgement

References

[1] *Abel, U. and Ivan, M.* Asymptotic expansion of the multivariate Bernstein polynomials on a simplex // *Approx. Theory Appl. N. S.*. 2000. V. 16, N 3. P. 85–93.

[2] *Bernstein, S. N.* Complement à l'article de E. Voronovskaja "Determination de la forme asymptotique de l'approximation des fonctions par les polynômes de S. Bernstein" // *Dokl. Akad. Nauk SSSR, Ser. A (Russian)*, 1932. V. 4. P. 86–92.

[3] *Bernstein, S. N.* Communications de la Société mathématique de Kharkow. 2-ée série, 1912, V. 13, N 1, P. 1–2. (See also http://topo.math.u-psud.fr/∼slc/TeX/lcs/legacy_math/bernstein/bernstein_doyle.pdf)

[4] *Chlodovsky, I. N.* On some properties of SN Bernstein polynomials // Moscow, Proc. 1st All-Union congress of mathematics (Russian), (Kharkov, 1930). P. 22.

[5] *Floater, M. S.* On the convergence of derivatives of Bernstein approximation // *J. Approx. Theory.* 2005. V. 134, N 1. P. 130–135.

[6] *Krylov, N. V. Introduction to the Theory of Random Processes* AMS, Providence, Rhode Island, 2002.

[7] *Lorentz, G. G.* Zur Theorie der Polynome von S. Bernstein // *Mathematical Collections* 1937. V. 2(44), No. 3. P. 543–556.

[8] *Lorentz, G. G. Bernstein Polynomials* / 2nd ed. New York: AMS Chelsea Publishing Co., 1986.

[9] *Pop, O. T.* Voronovskaja-type theorem for certain GBS operators // *Glas. Mat. Ser III.* 2008. V. 43(63), N 1. P. 179–194.

[10] *Popoviciu, T.* Sur l'approximation des fonctions convexes d'ordre supérieur // *Mathematica.* 1935. V. 10. P. 49–54.

[11] *Shiryaev, A. N Probability* New York, Springer, 2nd edition, 1995.

[12] *Telyakovskii, S. A.* On the rate of approximation of functions by the Bernstein polynomials // *Proc. Inst. Math. Mech.*, 2009. V. 264, suppl. 1. P. 177–184.

[13] *Tikhonov, I. V., Sherstyukov, V. B.* The module function approximation by Bernstein polynomials // *Bulletin of Chelyabinsk State University (Russian)* 2012. No. 26(15). P. 6-40.

[14] *Veretennikov, A. Yu., Veretennnikova, E. V.* On convergence of partial derivatives of multivariate Bernstein polynomials // *Mathematics, Informatics and Physics in Science and Education* / Collection of scientific papers for 140th anniversary of MSPU (Russian). Moscow, "Prometei", 2012. P. 39–42.

[15] *Voronovskaya, E. V.* Determination of the asymptotic form of approximation of functions by the Bernstein polynomials // *Dokl. Akad. Nauk SSSR, Ser. A (Russian)*, 1932. V. 4. P. 74–85.

[16] *Zorich, V. A. Mathematical Analysis II.* Berlin, Springer, 2004.