Bernstein Polynomials and Learning Theory

Dietrich Braess

Ruhr–Universität Bochum, Fakultät für Mathematik, D–44780 Bochum, Germany

Thomas Sauer

Justus–Liebig–Universität Gießen, Lehrstuhl für Numerische Mathematik, Heinrich–Buff–Ring 44, D–35392 Gießen, Germany

Abstract

When learning processes depend on samples but not on the order of the information in the sample, then the Bernoulli distribution is relevant and Bernstein polynomials enter into the analysis. We derive estimates of the approximation of the entropy function $x \log x$ that are sharper than the bounds from Voronovskaja's theorem. In this way we get the correct asymptotics for the Kullback-Leibler distance for an encoding problem.

Key words: Bernstein polynomials, entropy function, learning theory, optimal encoding
PACS:

1 Introduction

The approximation properties of the Bernstein polynomials for the *entropy* function

$$f(x) := -x \log x - (1-x) \log(1-x) \tag{1.1}$$

are of interest since $f''(x) = -[x(1-x)]^{-1}$ and, according to the Voronovskaja theorem, cf. [10, p. 22], the *pointwise* limit

$$\lim_{n\to\infty} n\{f(x) - B_n[f](x)\} = \frac{1}{2} \tag{1.2}$$

Email addresses: `braess@math.ruhr-uni-bochum.de` (Dietrich Braess), `tomas.sauer@math.uni-giessen.de` (Thomas Sauer).

is constant for all $x \in (0,1)$. On the other hand, the difference $f - B_n[f]$ assumes the value 0 at the boundary points $x = 0, 1$ for all $n \in \mathbb{N}$. Thus the convergence in (1.2) cannot be uniform, though f does belong to the O–*saturation class* for the Bernstein polynomials. Further information on the global approximation behavior can be found e.g. in [3,8,14]. More surprising, however, is the fact that although $f(x) - B_n[f](x) \geq 0$ for all $x \in [0,1]$ with equality exactly for $x = 0, 1$, the value $\frac{1}{2}$ is always exceeded by the approximation which can be phrased as

$$c := \liminf_{n\to\infty} \sup_{x\in[0,1]} n\left\{f(x) - B_n[f](x)\right\} > \frac{1}{2}.$$

It is worthwhile to mention that the abscissas where the maximum is assumed tend to the boundary like $O\left(n^{-1}\right)$ as n increases.

It can be shown relatively easily that

$$f(x) - B_n[f](x) \geq \frac{1}{2n} + o\left(n^{-1}\right)$$

holds uniformly in the interior of $[0,1]$ as long as boundary regions of size $O\left(n^{-1+\epsilon}\right)$, $\epsilon > 0$, are excluded, see (2.7). This, however, is insufficient for the application we bear in mind, in particular as this way the points where the maximal deviation takes place are not captured. For that reason we establish an improved estimate in Theorem 1 which extends up to boundary regions of size $O\left(n^{-1}\right)$. The crucial tool to achieve this goal is a one–sided estimate for the Bernstein polynomials of convex functions which is applicable to the Taylor polynomials of f here.

The improved estimate enables us to close a gap in an application from Learning Theory which is concerned with the optimal encoding of the output of a random source under the assumption that a sample of length n is available. Carefully analyzing the difference between the entropy function and its approximating Bernstein polynomials, we obtain improved asymptotics compared to [9]. For points close the to left boundary, i.e., when nx is (uniformly) bounded, the asymptotical behavior is captured by a function that can be accessed numerically. In contrast to [9] those numerical estimates are only needed for the representation of one particular *univariate* function. The gap for x between cn^{-1} and $cn^{-1-\epsilon}$ which had still been present in [5] is thus closed. It is remarkable that the improved asymptotical estimate becomes available due to methods from Approximation Theory which investigate the approximation behavior of $n\left\{f - B_n[f]\right\}$, thus remaining in the "finite" Bernoulli probability distribution, instead of passing to the Poisson distribution as it is traditional in Bayesian statistics.

The univariate entropy function f from (1.2) corresponds to sources that use a binary alphabet consisting of *two* symbols only. To study more general sources, we need to extend the estimates to multivariate Bernstein polynomials on simplices which will be done by reducing it to sums over univariate Bernstein polynomial approximations.

2 Interior Estimate

In this section we consider the approximation of the univariate function (1.1) by Bernstein polynomials

$$B_n[f](x) := \sum_{k=0}^{n} B_k^n(x) f\Big(\frac{k}{n}\Big)$$

where

$$B_k^n(x) := \binom{n}{k} x^k (1-x)^{n-k}. \tag{2.1}$$

The aim of this section is an estimate of the approximation error in the interior that is sharper than Voronovskaja's bound.

Theorem 1 *Let f be defined by (1.1). Then*

$$f(x) - B_n[f](x) \geq \frac{1}{2n} + \frac{1}{20n^2 x(1-x)} - \frac{1}{12n^2} \qquad \text{for } \frac{15}{n} \leq x \leq 1 - \frac{15}{n}. \tag{2.2}$$

The following observation from [2,11] provides a useful tool, and its short proof will be given for the sake of completeness.

Lemma 2 *If f is concave in $(0,1)$, then we have*

$$f(x) - B_n[f](x) \geq 0 \quad \text{for} \ \ 0 \leq x \leq 1.$$

PROOF. Given $x_1 \in [0,1]$, let Q_1 be the linear polynomial that interpolates f and f' at x_1. Since f is concave, we have $Q_1 - f \geq 0$ in $[0,1]$. The mapping $f \mapsto B_n[f]$ is performed by a positive linear operator, and it follows that $B_n[Q_1 - f] \geq 0$. Moreover, linear functions are reproduced by Bernstein polynomials. Hence,

$$
\begin{aligned}
B_n[f](x_1) &= B_n[f-Q_1](x_1) + B_n[Q_1](x_1) \le B_n[Q_1](x_1) = Q_1(x_1) \\
&= f(x_1).
\end{aligned}
$$

This holds for any $x_1 \in [0,1]$, and the proof is complete. $\square$

A direct consequence is the following.

Corollary 3 *Let $n \ge 4$ be an even number, $f^{(n)} \le 0$ in $(0,1)$ and Q_{n-1} be the Taylor polynomial of degree n to f at some x_1 in $(0,1)$. Then we have in $[0,1]$:*

$$
f - B_n[f] \ge Q_{n-1} - B_n[Q_{n-1}].
$$

The inequality follows immediately from Lemma 2. The second derivative of $f - Q_{n-1}$ vanishes at x_1 and by Taylor's formula it is not positive in the interval.

The evaluation of the Bernstein polynomials of Taylor's polynomials requires the following expressions

$$
B_n[(x-x_0)^s](x_0) = \sum_{k=0}^{n} B_k^n(x_0)\Big(\frac{k}{n} - x_0\Big)^s.
$$

The right hand side coincides with $n^{-s}T_{ns}(x_0)$ in terms of the functions defined in [10, p. 13] and are provided up to $s = 5$ in [10, p. 14].

Proposition 4 *Let $0 \le x_0 \le 1$. Then we have $x = x_0$,*

$$
\begin{aligned}
B_n[(x-x_0)^2] &= \frac{x(1-x)}{n}, \\
B_n[(x-x_0)^3] &= \frac{x(1-x)}{n^2}(1-2x), \\
B_n[(x-x_0)^4] &= 3\,\frac{x^2(1-x)^2}{n^2} + \frac{x(1-x)}{n^3}[1-6x(1-x)], \\
B_n[(x-x_0)^5] &= \Big\{10\frac{x^2(1-x)^2}{n^3} + \frac{x(1-x)}{n^4}[1-12x(1-x)]\Big\}(1-2x).
\end{aligned}
$$

The monomials $(x-x_0)^m, m > 2$, cause only contributions of the order n^{-2}. This is consistent with Voronovskaja's theorem.

We split the symmetric entropy function (1.1) into two parts in view of the generalization to the higher dimensional case

$$
\begin{aligned}
&f(x) = g(x) + g(1-x),\\
&g(x) := -x \log x.
\end{aligned}
\tag{2.3}
$$

Obviously,

$$
\begin{aligned}
g'(x) &= -\log x - 1,\\
g^{(k)}(x) &= (-1)^{k+1}\frac{(k-2)!}{x^{k-1}} \qquad \text{for} \quad k > 1,
\end{aligned}
$$

and $g^{(2m)} < 0$. Taylor's polynomial of degree 5 has the form

$$
Q_5(x) = \sum_{k=2}^{5} \frac{(-1)^{k-1}}{k(k-1)\,x_0^{k-1}}(x-x_0)^k + \text{ linear polynomial.}
$$

The Bernstein polynomial for Q_5 is now evaluated at $x = x_0$ by using Lemma 4,

$$
\begin{aligned}
Q_5(x) - B_n\left[Q_5\right](x) &= \frac{1-x}{2n} - \frac{(1-x)(1-2x)}{6n^2 x} \qquad (2.4)\\
&+\frac{(1-x)^2}{4\,n^2 x} + \frac{(1-x)}{12n^3x^2}\left[1-6x(1-x)\right]\\
&-\frac{(1-x)^2(1-2x)}{2n^3x^2} - \frac{(1-x)(1-2x)}{20\,n^4x^3}\left[1-12x(1-x)\right]\\
&= (1-x)\left[\frac{1}{2n} + \frac{1+x}{12n^2x}\right]\\
&+\frac{1}{2n^3x^2}\left[\frac{(1-x)\,(1-6x(1-x))}{6} - (1-x)^2(1-2x)\right]\\
&-\frac{(1-x)(1-2x)}{20\,n^4x^3}\left[1-12x(1-x)\right]\\
&=: \frac{1-x}{2n} + \frac{1}{12n^2x} - \frac{x}{12n^2} + \frac{1}{2n^3x^2}\,R_1(x) + \frac{1}{20\,n^4x^3}\,R_2(x). \qquad (2.5)
\end{aligned}
$$

Next we estimate the function R_1 in the term of order n^{-3} in (2.5)

$$
\begin{aligned}
R_1(x) &= (1-x)\left(\tfrac{1}{6} - x(1-x) - (1-x)(1-2x)\right)\\
&= (1-x)\left(\tfrac{1}{6} - (1-x)^2\right) \geq (1-x)\left(\tfrac{1}{6} - 1\right) \geq -\tfrac{5}{6}.
\end{aligned}
$$

The function R_2 will estimated from below by the trivial bound $R_2(x) \geq -2$. Hence,

$$
\frac{R_1(x)}{2n^3x^2} + \frac{R_2(x)}{20\,n^4x^3} \geq -\frac{1}{n^2x}\left(\frac{5}{12nx} + \frac{1}{10(nx)^2}\right).
$$

We estimate now all the terms in (2.4) with a singularity at zero for $nx \geq 6$:

$$\frac{1}{12n^2x} + \frac{R_1(x)}{2n^3x^2} + \frac{R_2(x)}{20\,n^4x^3} \geq \frac{1}{n^2x}\left(\frac{1}{12} - \frac{5}{12\,nx} - \frac{1}{60\,nx}\right) \geq \frac{1}{12\,n^2x}\left(1 - \frac{6}{nx}\right).$$

By collecting terms we obtain.

Theorem 5 *Let g be defined by (2.3). For $x \geq 15/n$ we have*

$$\begin{aligned} g(x) - B_n[g](x) &\geq \frac{1-x}{2n} + \frac{1}{12\,n^2x}\left(1 - \frac{6}{nx}\right) - \frac{x}{12n^2} \\ &\geq \frac{1-x}{2n} + \frac{1}{20n^2x} - \frac{x}{12n^2}. \end{aligned} \tag{2.6}$$

A symmetry argument yields the corresponding estimate for $g(1-x)$, and the proof of Theorem 1 is also complete.

Obviously, an estimate is more easily determined if it is based only on Taylor's polynomial of degree 3 and only the first line of (2.4) is taken into account. We note that the resulting estimate

$$f(x) - B_n[f]x \geq \frac{1}{2n} - \frac{(1-2x)^2}{6n^2x(1-x)}, \tag{2.7}$$

however, is not sufficient for our purpose.

3 Behavior at the Boundary

In Theorem 1 the neighborhood of the boundary of the interval is excluded. The behavior near the (left) boundary of the interval will be described by a function of the variable

$$z = nx. \tag{3.1}$$

The function will be given by a power series; cf. Krichevskiy [9], but the required properties will be determined by numerical computations.

Recalling (2.3) we set

$$\mathcal{L}_n(z) := n\{g - B_n[g]\}\left(\frac{z}{n}\right).$$

The handling of the Binomial coefficients will be simplified by the notation; cf. [1]

$$n^{\underline{k}} := n(n-1)(n-2)\dots(n-k+1). \tag{3.2}$$

Since the linear function $x \log n$ is reproduced by Bernstein polynomials [10], it follows that

$$\begin{aligned}
\mathcal{L}_n(z) &= -nx\log x + n\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\frac{k}{n}\log\frac{k}{n} \\
&= -nx\log x - nx\log n + n\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\Big[\frac{k}{n}\log\frac{k}{n}+\frac{k}{n}\log n\Big] \\
&= -z\log z + \sum_{k=1}^{n}\binom{n}{k}\Big(\frac{z}{n}\Big)^k(1-\frac{z}{n})^{n-k}k\log k \\
&= -z\log z + \Big(1-\frac{z}{n}\Big)^n\sum_{k=2}^{n}\frac{n^{\underline{k}}}{n^k}\frac{1}{k!}\Big[z\Big(1-\frac{z}{n}\Big)^{-1}\Big]^k k\log k.
\end{aligned}$$

Obviously $z\big(1-\frac{z}{n}\big)^{-1}$ converges uniformly to z on the compact interval $[0,15]$. Moreover, $\frac{n^{\underline{k}}}{n^k} \to 1$ for each k. The coefficients of the power series converge for $n\to\infty$, and the limits decrease as fast as the coefficients of the exponential function. Therefore, by extending the well-known argument for uniform convergence of power series we conclude that $\mathcal{L}_n$ converges uniformly on the interval $[0,15]$ to

$$\begin{aligned}
\mathcal{L}(z) &:= -z\log z + e^{-z}\sum_{k=2}^{\infty}\frac{z^k}{k!}k\log k \\
&= -z\log z + ze^{-z}\sum_{k=1}^{\infty}\frac{z^k}{k!}\log(k+1).
\end{aligned} \tag{3.3}$$

The complementary function from (2.3)

$$h(x) := g(1-x) = -(1-x)\log(1-x) = x - \sum_{k=2}^{\infty}\frac{x^k}{k(k-1)}$$

and the difference to its Bernstein polynomial are easily estimated by

$$\begin{aligned}
|h(x) - B_n[h](x)| &\le |h(x)-x| + |B_n[h-x](x)| \\
&\le x^2 + x^2 + \frac{x}{n} \le \frac{465}{n^2} \qquad \text{for } \ 0\le x\le\frac{15}{n}.
\end{aligned} \tag{3.4}$$

Therefore, $g(1-x)$ does not contribute to the asymptotics and

$$\lim_{n\to\infty} n\{f - B_n[f]\}\Big(\frac{z}{n}\Big) = \mathcal{L}(z) = -z\log z + ze^{-z}\sum_{k=1}^{\infty}\frac{z^k}{k!}\log(k+1).$$

The function $\mathcal{L}(z)$ is depicted in Figure 1.

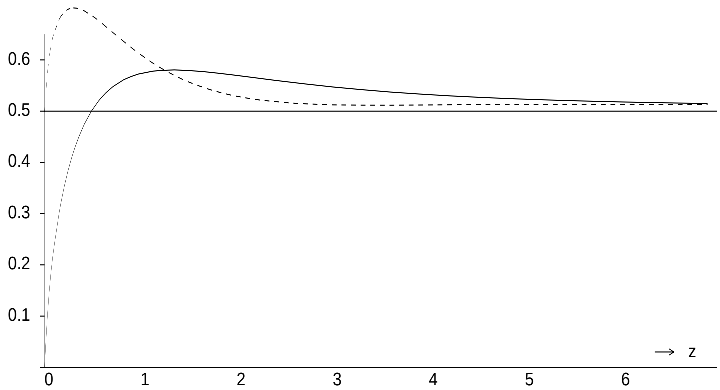

Fig. 1. Asymptotics of the error at the left end of the interval for $-x\log x$*:* $\mathcal{L}(z)$ *and the modification in (3.6), (dashed).*

The descent of $f - B_n[f]$ from Voronovskaja's bound to zero is confined to an intervals of length 1. This is expressed in the form

$$\lim_{n\to\infty}\min_{0\le x\le 1} n\Big\{f(x) - B_n[f](x) + \frac{1}{2n}[B_0^n(x) + B_n^n(x)]\Big\} = \frac{1}{2}. \tag{3.5}$$

This remains true if in addition the non-negative function $\frac{0.4}{n}[B_1^n + B_{n-1}^n]$ is subtracted from $\mathcal{L}$. To verify this, we have depicted in Figure 1

$$\mathcal{L}(z) + e^{-z}\left(\frac{1}{2} - \frac{2}{5}z\right) \tag{3.6}$$

in addition to $\mathcal{L}$. Summing up, the asymptotics of $n(f - B_n[f])$ is now completely characterized by Theorem 1 and (3.5) or Figure 1, respectively.

4 Extension to several variables

We are now turning our attention to the approximation behavior of Bernstein polynomials for the *multivariate* entropy function

$$f(u) = -\sum_{j=0}^{m} u_j \log u_j, \tag{4.1}$$

where the components u_j, $j = 0, \ldots, m$ of the vector

$$u \in \Delta_m := \left\{ u = (u_0, \ldots, u_m) \in \mathbb{R}^{m+1} \;:\; u_j \geq 0, \sum_{j=0}^{m} u_j = 1 \right\},$$

can be viewed either as probabilities or as *barycentric coordinates* in the *unit simplex* Δ_m. The multivariate Bernstein polynomial on the simplex now takes the form

$$B_n[f](u) = \sum_{\alpha \in \mathcal{K}_{m,n}} f\left(\frac{\alpha}{n}\right) B_\alpha(u),$$
$$B_\alpha(u) := \binom{n}{\alpha} u^\alpha = \frac{n!}{\alpha_0! \cdot \alpha_m!} u_0^{\alpha_0} \cdots u_m^{\alpha_m},$$

where

$$\mathcal{K}_{m,n} := \left\{ \alpha = (\alpha_0, \ldots, \alpha_m) \in \mathbb{N}_0^{m+1} \;:\; n = |\alpha| := \sum_{j=0}^{m} \alpha_j \right\}$$

is the set of all *homogeneous* multiindices of *length* $|\alpha| = n$. Note that in the above notation the univariate basis polynomial $B_k^n(x)$ coincides with $B_{n-k,k}(1-x, x)$.

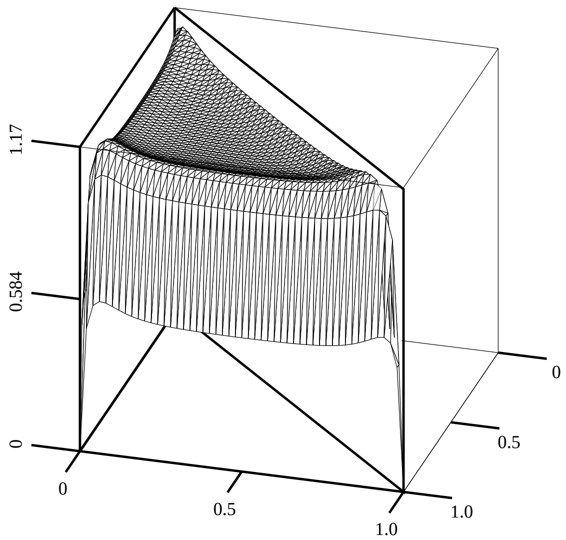

Fig. 2. Approximation behavior of $n\{f - B_n[f]\}$, *here for* $n = 20$ *and* $m = 2$.

Before we investigate the multivariate approximation behavior of $B_n[f]$, we depict the error of approximation $f - B_n[f]$ in Figure 2, showing that again at the boundary, in particular close to the corners, the error of approximation is significantly higher than the “plateau” of value 1 that is approached in the interior of the triangle. Also note that at the boundary the *univariate* approximation behavior is visible, now however with the associated limit value $\frac{1}{2}$.

In the following lemma we will state an elementary identity which will allow us to reduce the approximation of the function f from (4.1) to the univariate case.

Lemma 6 *For any set of functions* $G_j : [0,1] \times \mathbb{N}_0 \to \mathbb{R}$ *we have*

$$\sum_{\alpha \in \mathcal{K}_{m,n}} B_\alpha(u) \sum_{j=0}^{m} G_j(u_j, \alpha_j) = \sum_{j=0}^{m} \sum_{k=0}^{n} B_k^n(u_j)\, G_j(u_j, k). \tag{4.2}$$

In particular, if

$$g = \sum_{j=0}^{m} g_j(u_j), \qquad \text{then} \qquad B_n[g](u) = \sum_{j=0}^{m} B_n[g_j](u_j)\ . \tag{4.3}$$

PROOF. Define, for any $\alpha \in \mathcal{K}_{m,n}$ and $0 \le j \le m$ the reduced multiindex $\widehat{\alpha}^j := \alpha - \alpha_j e_j$, which coincides with α except that its j–th component has zero value. We decompose the basis polynomials B_α in a fashion similar to tensor products, cf. [4,12], to obtain

$$\begin{aligned}
&\sum_{\alpha \in \mathcal{K}_{m,n}} \sum_{j=0}^{m} G_j(u_j, \alpha_j)\ B_\alpha(u) \\
&= \sum_{j=0}^{m} \sum_{\alpha_j=0}^{n} \sum_{\widehat{\alpha}^j \in \mathcal{K}_{m-1,n-\alpha_j}} G_j(u_j, \alpha_j)\ B_\alpha(u) \\
&= \sum_{j=0}^{m} \sum_{\alpha_j=0}^{n} G_j(u_j, \alpha_j) \frac{n!}{(n-\alpha_j)!\alpha_j!} u_j^{\alpha_j} \times \\
&\quad \times \sum_{\widehat{\alpha}^j \in \mathcal{K}_{m-1,n-\alpha_j}} \frac{(n-\alpha_j)!}{\alpha_0! \cdots \alpha_{j-1}! \alpha_{j+1}! \cdots \alpha_m!} u^{\widehat{\alpha}^j} \\
&= \sum_{j=0}^{m} \sum_{\alpha_j=0}^{n} G_j(u_j, \alpha_j)\, u_j^{\alpha_j} \binom{n}{\alpha_j} (u_0 + \cdots + u_{j-1} + u_{j+1} + \cdots + u_m)^{n-\alpha^j} \\
&= \sum_{j=0}^{m} \sum_{k=0}^{n} G_j(u_j, k) \binom{n}{k} u_j^k\ (1-u_j)^{n-k} = \sum_{j=0}^{m} \sum_{k=0}^{n} G_j(u_j, k)\, B_k^n(u_j)\ .
\end{aligned}$$

This proves (4.2). By setting $G_j(u_j, \alpha_j) := g_j(k/n)$ we obtain (4.3). $\square$

Combining Lemma 6 with Theorem 5, we derive the multivariate counterpart of (2.2).

Theorem 7 *Let the function f be defined in (4.1). For any $u \in \Delta_m$ such that $u_j \geq 15/n$ for $j = 0, 1, \dots, m$ we have*

$$f(u) - B_n[f](u) \geq \frac{m}{2n} + \frac{1}{20\, n^2} \sum_{j=0}^{m} \frac{1}{u_j} - \frac{1}{12n^2}. \tag{4.4}$$

To extend our estimate to the boundary also in the multivariate case, we will again appeal to (4.3) and make use of the univariate estimates obtained in the preceding chapter. To that end, we define for $u \in \Delta_m$ the two index sets

$$I_{\geq} := \left\{ j \ : \ u_j \geq \frac{15}{n} \right\} \qquad \text{and} \qquad I_{<} := \left\{ j \ : \ u_j < \frac{15}{n} \right\}$$

of cardinality $m+1-k$ and $k := \# I_{<}$, respectively. Then we obtain from (2.6) that

$$\begin{aligned} n \sum_{j \in I_{\geq}} \{g - B_n[g]\}(u_j) &\geq \sum_{j \in I_{\geq}} \left(\frac{1-u_j}{2} + \frac{1}{20 n u_j} - \frac{u_j}{12n} \right) \\ &\geq \frac{m+1-k}{2} - \left(\frac{1}{2} + \frac{1}{12n} \right) \sum_{j \in I_{\geq}} u_j \\ &\geq \frac{m+1-k}{2} - \left(\frac{1}{2} + \frac{1}{12n} \right) \left(1 - \frac{12k}{n} \right) = \frac{m-k}{2} + O\left(n^{-1}\right). \end{aligned}$$

Taking also into account (3.3) we thus end up with

$$n \{f - B_n[f]\}(u) \geq \frac{m-k}{2} + \sum_{j \in I_{<}} \mathcal{L}(n u_j) + O\left(n^{-1}\right),$$

or, with any $z \in \mathbb{R}^{n+1}$ such that $|z| = n$,

$$n \{f - B_n[f]\}\left(\frac{z}{n}\right) \geq \frac{m-k}{2} + \sum_{j \,:\, z_j \leq 15} \mathcal{L}(z_j) + O\left(n^{-1}\right). \tag{4.5}$$

Since the basis polynomials B_{e_j} decrease exponentially away from the vertices of the simplex, now the counterpart of (3.5) becomes

$$\lim_{n \to \infty} \min_{u \in \Delta_m} n \left\{ f(u) - B_n[f](u) + \frac{1}{2n} \sum_{j=0}^{m} B_{e_j}(u) \right\} = \frac{1}{2}, \tag{4.6}$$

which is always assumed on the boundary, see Figure 2. In general, on a k–dimensional face of the boundary the minimum on the interior of it would be $\frac{k}{2}$.

5 An Application from Learning Theory

Theorem 1 and the knowledge of the Bernstein approximation in the region next to the boundary enables us to determine the exact asymptotics for a problem in learning theory.

The symbols $A_0, A_1, \ldots, A_m$ of an alphabet with $m+1$ letters are to be encoded. The length of the codes may be different for the letters. Following [13] it is possible to have a code with length $\log \frac{1}{q_i}$ for the letter A_i if $\sum_{i=0}^m q_i = 1$. If the symbol A_i is found with the probability p_i, the expectation value of the code length is

$$\sum_{i=0}^m p_i \log \frac{1}{q_i}.$$

The minimum of this expression is attained if $q_i = p_i$ for all i. If the lengths q_i differ from the optimal values, there is a redundancy, i.e. a difference to the minimum of

$$\sum_{i=0}^m p_i \log \frac{p_i}{q_i}. \tag{5.1}$$

First we restrict our attention to the special case $m = 1$. Here the sum (5.1) may be rewritten as

$$p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}\,, \tag{5.2}$$

if we write $p_1 = p$, $p_0 = 1-p$, $q_1 = q$, and $q_0 = 1-q$.

The probability p is unknown, but we have got a sample with n letters. The encoding will be performed on the base of the information, how often the symbol A_1 is contained in the sample. Due to Bernoulli, the probability for finding it k times in the sample is

$$\binom{n}{k}(1-p)^{n-k}p^k = B_k^n(p).$$

Now, an appropriate rule $k \mapsto Q(k), 0 \le k \le n$, is to be found for the encoding procedure. If the sample contains the symbol A_1 exactly k times, the encoding for the parameter $q_k = Q(k)$ is chosen. The associated contribution to the redundancy is

$$B_k^n(p)\Big[p \log \frac{p}{q_k} + (1-p) \log \frac{1-p}{1-q_k}\Big].$$

Summing over all k we obtain the expectation value of the redundancy [7,9]. Therefore, the following problem arises:

Find numbers $q_k \in (0,1),\ k = 0, 1, \ldots, n$, *such that the worst case redundancy*

$$\sup_{0 \le x \le 1} F_n(x) \tag{5.3}$$

is minimized where

$$\begin{aligned} F_n(x) &:= \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \Big[x \log \frac{x}{q_k} + (1-x) \log \frac{1-x}{1-q_k}\Big] \qquad (5.4) \\ &= \sum_{k=0}^{n} B_k^n(x) \Big[x \log \frac{x}{q_k} + (1-x) \log \frac{1-x}{1-q_k}\Big]. \end{aligned}$$

The challenge consists of determining the numbers q_k in such a way that the optimal asymptotics are obtained inside the interval and on its boundary simultaneously.

There are already several results for the rule called *add-β rule*,

$$q_k^\beta := \frac{k+\beta}{n+2\beta} \quad \text{for} \quad k = 0, 1 \ldots, n. \tag{5.5}$$

The parameter β describes the deviation of q_k from k/n, i.e. from the relative frequency of A_1 in the sample. In particular, the add-one rule is called *Laplace's rule of succession*, and the add-half rule is *Jeffreys' rule.* Krichevskiy [9] reported that $\beta_0 = 0.50922$ leads to

$$\lim_{n \to \infty} \sup_{0 \le x \le 1} n F_n^{\beta_0}(x) = \beta_0 = 0.50922,$$

while the corresponding number for Jeffrey's prior is 0.5106 due to our calculations. Moreover, Cover [6] had shown the lower bound

$$\lim_{n \to \infty} \sup_{0 \le x \le 1} n F_n(x) \ge 0.5 \quad \text{for all choices of } q\,,$$

by applying a suitable functional and the add-one rule.

We will close the gap by applying Theorem 1; cf. [5]. Our point of departure is the add-3/4 rule. This rule is optimal in the interior, and it will be modified later to cover also the subdomain next to the boundary. — Since we fix the parameters in the one-dimensional case such that $q_k + q_{n-k} = 1$, it follows that $F(1-x) = F(x)$ and

$$F_n(x) = G_n(x) + G_n(1-x)$$

where

$$G_n(x) := \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} x \log \frac{x}{q_k}\,. \tag{5.6}$$

Following Forster and Warmuth [7] we study the function for $n-1$ letters and separate the Bernstein polynomial for the function g defined in (2.3)

$$\begin{aligned} G_{n-1}(x) &= -\sum_{k=0}^{n-1} \binom{n-1}{k} x^{k+1}(1-x)^{n-1-k} \log q_k - g(x) \\ &= -\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \frac{k}{n} \log q_{k-1} - g(x) \\ &= \frac{1}{n} \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} k \log \frac{k/n}{q_{k-1}} + \{B_n[g](x) - g(x)\}. \end{aligned} \tag{5.7}$$

The add-β rule (5.5) for $n-1$ reads $q_k = \frac{k+\beta}{n-1+2\beta}$. We can extract the terms with $\frac{k}{n} \log \frac{n-1+2\beta}{n}$ since they belong to a Bernstein polynomial for a linear function that is reproduced,

$$\begin{aligned} G^{\beta}_{n-1}(x) &= \frac{1}{n} \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} k \log \frac{k}{k-1+\beta} + x \log \frac{n-1+2\beta}{n} \\ &\quad + B_n[g](x) - g(x). \end{aligned} \tag{5.8}$$

Now we fix $\beta = \beta_* = 3/4$, since the following estimate of the power series of the logarithm is an helpful upper bound only for $3/4 \le \beta \le 1$,

$$\begin{aligned} k \log \tfrac{k}{k-1/4} &= -k \log(1 - \tfrac{1}{4k}) = k[\tfrac{1}{4k} + \tfrac{1}{2(4k)^2} + \ldots] \\ &\le \frac{1}{4} + \frac{1}{32k} + \frac{4}{3}\frac{1}{192k^2} \\ &\le \frac{1}{4} + \frac{1}{32(k+1)} + \frac{1}{7(k+1)(k+2)}. \end{aligned}$$

When estimating the sum in (5.8) we note that $\binom{n}{k}\frac{x^k}{k+1} = x^{-1}\binom{n+1}{k+1}\frac{x^{k+1}}{n+1}$ and an analogous formula holds for the term with the quadratic denominator. Hence,

$$\begin{aligned}
&\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}k\log\frac{k}{k-1/4} \\
&\leq \sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}[\frac{1}{4}+\frac{1}{32(k+1)}+\frac{1}{7(k+1)(k+2)}] \\
&\leq \frac{1}{4}+\frac{1}{32(n+1)x}+\frac{1}{7(n+1)(n+2)x^2} \qquad (5.9)\\
&\leq \frac{1}{4}+\frac{1}{24(n+1)x} \qquad\qquad \text{for } x \geq \frac{15}{n}.
\end{aligned}$$

From the previous bound and Theorem 5 it follows that (for $x \geq \frac{15}{n}$)

$$\begin{aligned}
G_{n-1}^{\beta_*}(x) \leq& \frac{1}{n}\Big(\frac{1}{4}+\frac{1}{24(n+1)x}\Big)+x\log(1+\frac{1}{2n}) \\
&-\frac{1}{n}\Big(\frac{1-x}{2}+\frac{1}{20nx}-\frac{x}{12n}\Big) \\
\leq& \frac{1}{n}\Big(-\frac{1}{4}+x+\frac{x}{12n}\Big). \qquad (5.10)
\end{aligned}$$

Hence,

$$F_{n-1}^{\beta_*}(x) \leq \frac{1}{2n}+\frac{1}{12n^2} \qquad \text{for } \frac{15}{n} \leq x \leq 1-\frac{15}{n}. \qquad (5.11)$$

This is the required bound for the interior of the interval.

We note that we can drop the restriction $\frac{3}{4} \leq \beta \leq 1$ if we are interested in G_{n-1}^{β} for $x \geq 1/2$. In this case it is no drawback to have a larger factor in the $1/\left(n^2x^2\right)$ term in (5.9). Repeating the calculations with the power series of the logarithm, we obtain

$$G_{n-1}^{\beta}(x) \leq \frac{1}{n}\Big(\frac{1-x}{2}+\beta(2x-1)+\frac{3}{2n}\Big) \quad \text{for} \;\; x \geq \frac{1}{2},\; \frac{1}{2} \leq \beta \leq 1. \quad (5.12)$$

Now we consider the add-β rule at the left-hand boundary with emphasis on the choice $\beta = \beta_* = 3/4$. This will be done in the spirit and with the tools of Section 3. Since the bound (5.12) is sharp for $x \to 1$, [or alternately from (5.6)] it follows that

$$nG_{n-1}^{\beta}(1-z/n) \to \beta \qquad \text{for} \;\; n \to \infty. \qquad (5.13)$$

Moreover, terms of the size $xO(n^{-1})$ are obviously $zO(n^{-2})$. Now it follows from (5.8) with the argument as for the verification of (3.3); cf. Krichevskiy [9] that

$$\begin{aligned}\Phi^{\beta}(z) &:= \lim_{n\to\infty} nF^{\beta}_{n-1}\Big(\frac{z}{n}\Big)\\ &= e^{-z}\sum_{k=0}^{\infty}\frac{z^k}{k!}\,k\log\frac{k}{k-1+\beta}+\beta-\mathcal{L}(z)\\ &= \beta+z\log z-e^{-z}\sum_{k=1}^{\infty}\frac{z^k}{k!}\,k\log\Big(k-1+\beta\Big)\\ &= \beta+z\log z-ze^{-z}\sum_{k=0}^{\infty}\frac{z^k}{k!}\log\Big(k+\beta\Big). \end{aligned}\tag{5.14}$$

The convergence is uniform on compact intervals.

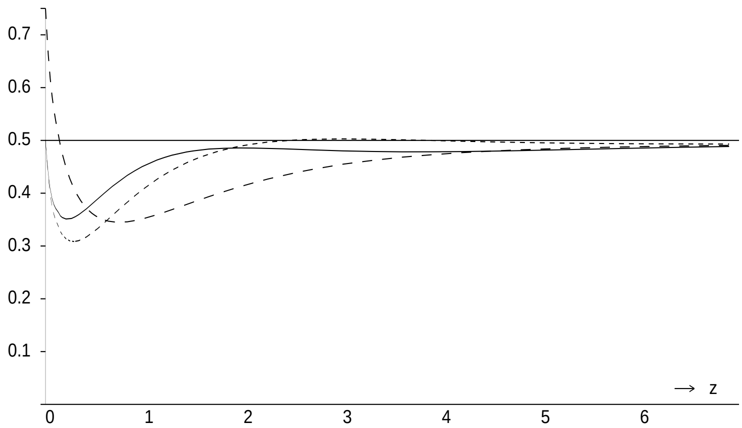

Fig. 3. Asymptotics of the entropy at the left end of the interval: $\tilde{\Phi}(z)$ and $\Phi^{\beta_*}(z)$. The maximal value is ≈ 0.5027 if only q_0 is modified.

Obviously, $\Phi^{\beta_*}(0) = \beta_* = 3/4$, and the choice (5.5) is not appropriate for the boundary region as depicted in Figure 3. The rule will become optimal if we modify q_0 and q_1 in an appropriate way. The symmetry requirement $q_k + q_{n-k} = 1$ causes that q_n and q_{n-1} will be also modified. Since we consider only the subdomain where $x \le \frac{1}{2}$, the contribution of q_n and q_{n-1} is $O(n^{-2})$. By taking differences we obtain

$$\begin{aligned}F_n(x)\; - F^{\beta}_n(x) &= (1-x)^n\,\Big\{x\log\tfrac{q^{\beta}_0}{q_0}+(1-x)\log\tfrac{1-q^{\beta}_0}{1-q_0}\Big\}\\ &\quad + nx(1-x)^{n-1}\Big\{x\log\tfrac{q^{\beta}_1}{q_1}+(1-x)\log\tfrac{1-q^{\beta}_1}{1-q_1}\Big\}+O(2^{-n})\ .\end{aligned}\tag{5.15}$$

The shifts will be bounded by $1/n$. Hence,

$$\log \frac{1-q_k^\beta}{1-q_k} = \log\{1 + \frac{q_k - q_k^\beta}{1-q_k}\} = q_k - q_k^\beta + O(n^{-2}) \quad \text{for} \quad k = 0,1. \tag{5.16}$$

We insert (5.16) into (5.15) to obtain

$$\begin{aligned} n[F_n(\tfrac{z}{n}) - F_n^\beta(\tfrac{z}{n})] &= e^{-z}\{z \log \tfrac{q_0^\beta}{q_0} + n(q_0 - q_0^\beta)\} \\ &\quad + z e^{-z}\{z \log \tfrac{q_1^\beta}{q_1} + n(q_1 - q_1^\beta)\} + O(n^{-1}). \end{aligned} \tag{5.17}$$

Now we are ready to present the final choice. Here, the add-β rules for $\beta = 1/2$, $3/4$ and 1 are combined,

$$q_k := \begin{cases} \frac{k+1/2}{n+5/4} & \text{if } k = 0\,, \\ \frac{k+1}{n+7/4} & \text{if } k = 1\,, \\ \frac{k+3/4}{n+7/4} & \text{if } k = n-1\,, \\ \frac{k+3/4}{n+5/4} & \text{if } k = n\,, \\ \frac{k+3/4}{n+3/2} & \text{otherwise}\ . \end{cases} \tag{5.18}$$

We insert the actual numbers in (5.17) and take the limit $n \to \infty$ to obtain

$$\tilde{\Phi}(z) = \Phi^{\beta_*}(z) + e^{-z}\Big[-\frac{1}{4} + z\{\log\frac{3}{2} + \frac{1}{4}\} - z^2 \log\frac{2}{7/4}\Big]. \tag{5.19}$$

The extra term in (5.19) is negative for $z > 5$ and does not spoil the bound from (5.10). Moreover, as is shown in Figure 3 for the interesting part of the interval, we have now

$$\tilde{\Phi}(z) \le \frac{1}{2} \qquad \text{for} \quad z \le 15.$$

This extends the estimate (5.11) to all of $[0,1]$ for the modified rule, and the upper bound of the asymptotics is complete for $m = 1$.

In view of the extension to the multivariate case $m > 1$ and an application of Theorem 7 we introduce the function

$$\tilde{G}_n(x) := G_n^{\beta_*}(x) + [-\frac{1}{4n} + x \log \tfrac{3}{2}]B_0^n(x) + [\frac{1}{4n} - x\log\tfrac{8}{7}]B_1^n(x) \tag{5.20}$$

which realizes the decomposition of F_n^* into barycentric coordinates: $F_n^*(x) = \tilde{G}_n(x) + \tilde{G}_n(1-x)$. Based on the preceding estimates we can immediately establish the inequality

$$\tilde{G}_n(x) \leq \frac{1}{n}(-\tfrac{1}{4} + x) + o(n^{-1}) \quad \text{for } \ 0 \leq x \leq 1, \tag{5.21}$$

where the $o(n^{-1})$ term is independent of x.

Indeed, if $x \geq \frac{15}{n}$, then (5.21) follows from $\tilde{G}_n(x) \leq G_n^{\beta_*}(x)$ and (5.10). If $x = \frac{z}{n} \leq \frac{15}{n}$, then recalling (5.13) we have $(n+1)\tilde{G}(z/n) \to \tilde{\Phi}(z) - \beta_* \leq -\frac{1}{4}$, and $x/n \leq 15/n^2$ together with the uniform convergence implies (5.21). Hence,

$$F_n^*(x) = \tilde{G}_n(x) + \tilde{G}_n(1-x) \leq \frac{1}{2n} + o\left(n^{-1}\right) ,$$

and we have verified the upper bound

$$\lim_{n\to\infty} \sup_{0\leq x\leq 1} nF_n^*(x) = \frac{1}{2} .$$

It cannot be improved due to the known lower bound [6].

The shift of q_0 was necessary since $\Phi^{\beta_*}(0) = \beta_* > \frac{1}{2}$, see also Figure 3. Therefore, q_0 is chosen according to Jeffrey's rule. This shift induces a deterioration in the interior, which can be compensated by a shift of q_1 into to opposite direction. From (5.19) we conclude that the additive term induced by the two shifts reduces the redundancy for $z = nx \geq 5$. As a consequence, the bound (5.11) is improved and not deteriorated. — This can be understood as motivation for *two* shifts.

6 Multivariate renormalization

We now turn our attention to alphabets consisting of the $m+1$ symbols $A_0, \dots, A_m$. Since the probabilities p_j of the symbols A_j are independent, the important information about a sample of the length n is how often each of the symbols appeared, which can be written as a multiindex $\alpha \in \mathcal{K}_{m,n}$. Now, a code with the parameter $q(\alpha) = (q_j(\alpha) \ : \ j = 0, \dots, m)$ is associated to any sample, and the deviation of the expected code length from entropy takes the form

$$\sum_{j=0}^{m} p_j \, \frac{p_j}{q_j(\alpha)}$$

as already mentioned above. Since the probability of α to appear is $B_\alpha(p)$, the *average deviation from entropy* is thus computed as

$$F_n(p) = F_{n,q}(p) = \sum_{\alpha\in\mathcal{K}_{m,n}} B_\alpha(p) \sum_{j=0}^{m} p_j \,\log \frac{p_j}{q_j(\alpha)}, \tag{6.1}$$

which is the natural generalization of (5.4), cf. [5]. To continue our approach from the theory of Bernstein polynomials, we will again use the symbol u for the probabilities p_j that are interpreted here as barycentric coordinates of an m-simplex.

To be able to apply Lemma 6 to (6.1) for a reduction to the univariate case, we would need that

$$q_j(\alpha) = q_j\left(\alpha_j\right), \qquad j = 0,\ldots,m, \quad \alpha \in \mathcal{K}_{m,n}, \tag{6.2}$$

an assumption that is too restrictive. The prediction rules we are going to use in the multivariate case will depend on both j and α as

$$q_j(\alpha) := \frac{\alpha_j + \beta(\alpha_j)}{n + \sum_{i=0}^{m} \beta(\alpha_i)}\,, \tag{6.3}$$

where

$$\beta(k) := \begin{cases} 1/2 & k = 0, \\ 1 & k = 1, \\ \beta_* = 3/4 & \text{otherwise.} \end{cases}$$

Note that these values do not have the property (6.2). For this reason we introduce renormalized quantities

$$\tilde{q}_j(\alpha) := r(\alpha_j) := \frac{\alpha_j + \beta(\alpha_j)}{n + (m+1)\beta_*}.$$

It will be no drawback that $\sum_j \tilde{q}_j(\alpha) \neq 1$ holds for some α. These auxiliary parameters can be accessed by Lemma 6 as follows:

$$F_{n,\tilde{q}}(u) = \sum_{\alpha\in\mathcal{K}_{m,n-1}} B_\alpha(u) \sum_{j=0}^{m} u_j \,\log \frac{u_j}{\tilde{q}_j(\alpha)} = \sum_{j=0}^{m}\sum_{k=0}^{n} u_j B_k^n(u_j) \log \frac{u_j}{r(k)}\,. \tag{6.4}$$

We consider the difference

$$\begin{aligned}\log \widetilde{q}_j(\alpha) - \log q_j(\alpha) &= \log \frac{n + \sum_{i=0}^m \beta(\alpha_i)}{n + (m+1)\beta_*} \\ &= \log \left(1 + \frac{\sum_{i=0}^m (\beta(\alpha_i) - \beta_*)}{n + (m+1)\beta_*} \right) \\ &\leq \frac{1}{n} \sum_{i=0}^m (\beta(\alpha_i) - \beta_*) + \frac{2m}{n^2} \sum_{i=0}^m |\beta(\alpha_i) - \beta_*| \,.\end{aligned}$$

For convenience, we will ignore the last sum since it contributes only to $O(n^{-2})$ and can be handled analogously to the first sum. Noting that this expression is independent of j we thus get

$$\begin{aligned}F_{n,q}(u) - F_{n,\tilde{q}}(u) &\leq \frac{1}{n} \sum_\alpha B_\alpha(u) \sum_{j=0}^m u_j \sum_{i=0}^m (\beta(\alpha_i) - \beta_*) \\ &= \frac{1}{n} \sum_\alpha B_\alpha(u) \sum_{i=0}^m (\beta(\alpha_i) - \beta_*) \\ &= \frac{1}{n} \sum_{i=0}^m \sum_{k=0}^m B_k^n(u_i)[\beta(k) - \beta_*] \\ &= \frac{1}{n} \sum_{i=0}^m [-\tfrac{1}{4} B_0^n(u_i) + \tfrac{1}{4} B_1^n(u_i)]. \end{aligned} \tag{6.5}$$

The summands in (6.5) correspond to two additive terms in (5.20). From the identities (6.4) and (6.5) it follows that

$$\begin{aligned}&F_{n,q}(u) = F_{n,\widetilde{q}}(u) + F_{n,q}(u) - F_{n,\tilde{q}}(u) \\ &= \sum_{j=0}^m \left[\sum_{k=0}^n B_k^n\,(u_j)\; u_j \log \frac{u_j}{\widetilde{\beta}(k)} - \frac{1}{4n} B_0^n\,(u_j) + \frac{1}{4n} B_1^n\,(u_j) \right] .\end{aligned} \tag{6.6}$$

The inner sum is now evaluated by comparing it with $G_n^{\beta_*}$,

$$\begin{aligned}&\sum_{k=0}^n B_k^n\,(u_j)\; u_j \log \frac{u_j}{r(k)} \\ &= \sum_{k=0}^n B_k^n(u_j) u_j \Big[\log \frac{u_j}{(k+\beta_*)/(n+2\beta_*)} + \log \frac{n + (m+1)\beta_*}{n + 2\beta_*}\Big] \\ &\quad + B_0^n(u_j) u_j \log \tfrac{3}{2} - B_1^n(u_j) u_j \log \tfrac{8}{7} \\ &= G_n^{\beta_*}(u_j) + \frac{(m-1)\beta_*}{n} u_j + o(n^{-1}) + B_0^n(u_j) u_j \log \tfrac{3}{2} - B_1^n(u_j) u_j \log \tfrac{8}{7} \,.\end{aligned}$$

Combining this with (6.4), (5.20) and (5.21) we obtain

$$
\begin{aligned}
F_{n,q}(u) &= \sum_{j=0}^{m}\Big[\tilde{G}_n(u_j) + \frac{m-1}{n}\frac{3}{4}u_j + o\left(n^{-1}\right)\Big] \\
&\le \sum_{j=0}^{m}\Big[\frac{1}{n}\Big(-\frac{1}{4} + u_j\Big) + \frac{m-1}{n}\frac{3}{4}u_j\Big] + o(n^{-1}) \\
&= \frac{1}{n}\Big[-\frac{m+1}{4} + 1 + \frac{3}{4}(m-1)\Big] + o(n^{-1}) \\
&= \frac{m}{2n} + o(n^{-1}).
\end{aligned}
$$

This completes the proof of our final result.

Theorem 8 *For the choice (6.3) of the prediction rule $q_j(\alpha)$ we get that*

$$\lim_{n\to\infty} \max_{u\in\Delta_m} n\, F_{n,q}(u) = \frac{m}{2},$$

which is the asymptotically optimal bound.

We remark that for the optimal q even the bound $(n+1)F_{n,q} \le \frac{m}{2}$ holds true, but of course passing from n to $n+1$ is irrelevant as far as asymptotics are concerned. To highlight the structure of $F_{n,q}$, we depict it for $m = 2$ and

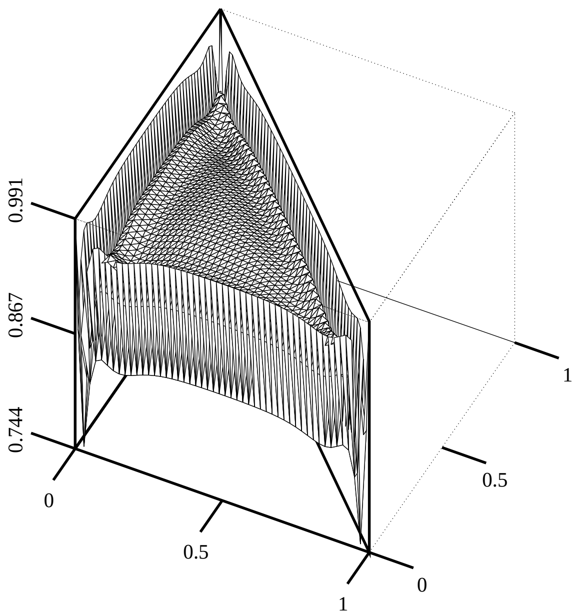

Fig. 4. The redundancy $(n+1)F_{n,q}$ for the optimal q from (6.3), $n = 25$.

$n = 25$ in Figure 4. Note that the extremal value is approached by the narrow ridges close to the boundary as well as in the interior of the simplex. It is also worthwhile to remark that along those boundary ridges the *univariate* behavior of the function can be observed. In fact, we always first have a sharp decrease from the value at the corners, which is due to the add-$\frac{1}{2}$ rules there, followed by a narrow "overshooting" due to the add-1 rule, then as small local minimum from which the function smoothly ascends to the interior limit function.

Acknowledgment. The authors want to thank H. Simon for bringing the problem to our knowledge and for many helpful discussions. We also thank J. Stöckler for some shorter proofs and the anonymous referee for carefully reading the manuscript.

References

[1] M. Aigner. Diskrete Mathematik. Vieweg, Braunschweig 1993

[2] V. G. Amel'kovič. A theorem converse to a theorem of Voronovskaya type. *Teor. Funkčiĭ, Funkcional Anal. i Prilozen, Vyp*, 2:67–74, 1966.

[3] H. Berens and G. G. Lorentz. Inverse theorems for Bernstein polynomials. *Indiana Univ. Math. J.*, 21:693–708, 1972.

[4] H. Berens and Y. Xu. K-moduli, moduli of smoothness, and Bernstein polynomials on a simplex. *Indag. Mathem.*, 2:411–421, 1991.

[5] D. Braess, J. Forster, T. Sauer, and H. U. Simon. How to achieve minimax expected Kullback-Leibler Distance from an unknown finite distribution. In *"Algorithmic Learning Theory", Proceedings of the 13th International Conference ALT 2002"* (N. Cesa-Bianchi, N. Numao, and R. Reischuk eds.), pp. 380–394, Springer-Verlag, Heidelberg 2002.

[6] T. M. Cover. Admissibility properties of Gilbert's encoding for unknown source probabilities. *IEEE Transactions on Information Theory*, 18:216–217, 1971.

[7] J. Forster and M. Warmuth. Relative expected instantaneous loss bounds. *Journal of Computer and System Sciences* 65:612–625, 2002.

[8] H.-P. Knoop and X. Zhou. The lower estimate for linear positive operators. II. *Result. Math.* 25: 315-330, 1994.

[9] R. E. Krichevskiy. Laplace's law of succession and universal encoding. *IEEE Transactions on Information Theory*, 44(1):296–303, 1998.

[10] G. G. Lorentz. Bernstein Polynomials. University of Toronto Press, 1953.

[11] T. Popoviciu. Sur l'approximation des fonctions convexes d'ordre superieur. *Mathematica*, 10:49–54, 1935.

[12] T. Sauer, The genuine Bernstein–Durrmeyer operator on a simplex. *Results in Mathematics*, 26:99–130, 1994.

[13] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[14] V. Totik. Approximation by Bernstein polynomials. *Amer. J. Math.* 116: 995-1018, 1994.