35

PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 9.1 (1988), p. 1-10

NONPARAMETRIC POLYNOMIAL DENSITY ESTIMATION

BY

Z. CIESIELSKI (SOPOT)

Abstract. A simple construction of polynomial estimators for densities and distributions on the unit interval is presented. For Lipschitz densities the error for the mean square deviation is characterized. The Casteljeau algorithm for calculating the values of the estimators is applied.

1. Introduction. The space of real polynomials of degree not exceeding m is denoted by Π_m. In Π_m we have the Bernstein basis, i.e.

$$\Pi_m = \operatorname{span}[N_{i,m},\ i = 0, \ldots, m],$$

where

$$N_{i,m}(x) = \binom{m}{i} x^i (1-x)^{m-i}, \qquad i = 0, \ldots, m.$$

The Casteljeau algorithm is based on the identity

$$N_{i,m}(x) = (1-x) N_{i,m-1}(x) + x N_{i-1,m-1}(x). \tag{1.1}$$

For given $w \in \Pi_m$

$$w(x) = \sum_{i=0}^{m} w_i N_{i,m}(x), \tag{1.2}$$

where the coefficients w_i are unique. Using (1.1) we find that, for $0 \leqslant k \leqslant m$,

$$w(x) = \sum_{i=0}^{m-k} w_i^{(k)}(x) N_{i,m-k}(x), \tag{1.3}$$

where $w_i^{(k)} \in \Pi_k$, and for $0 \leqslant k < m$ we have

$$w_i^{(k+1)}(x) = (1-x) w_i^{(k)}(x) + x w_{i+1}^{(k)}(x), \qquad i = 0, \ldots, m-k-1. \tag{1.4}$$

In particular, $w(x) = w_0^{(m)}(x) = \text{const}$.

Some more properties of the Bernstein polynomials will be needed. Our attention will be restricted to the interval $I = [0, 1]$ and the following notation will be used:

$$(f, g) = \int_I f(x) g(x) dx, \quad \|f\|_2 = \left(\int_I |f|^2\right)^{1/2}.$$

It is convenient to use simultaneously with $N_{i,m}$ the polynomials $M_{i,m} = (m+1) N_{i,m}$.

The following elementary properties of the polynomials $N_{i,m}$ and $M_{i,m}$ will be used:

1° $N_{i,m}(x) \geqslant 0$ for $x \in I$, $i = 0, \ldots, m$.

2° $\sum_{i=0}^{m} N_{i,m} = 1$.

3° $(M_{i,m}, 1) = 1$ for $i = 0, \ldots, m$.

4° For w as in (1.2) we have

$$Dw = m \sum_{i=0}^{m-1} \Delta w_i N_{i,m-1} = \sum_{i=0}^{m-1} \Delta w_i M_{i,m-1},$$

where $\Delta w_i = w_{i+1} - w_i$ and $Dw = dw/dx$.

5° For $i = 0, \ldots, m$ we have $DN_{i,m} = M_{i-1,m-1} - M_{i,m-1}$ with $M_{j,m} = 0$ whenever $j < 0$ or $j > m$.

2. Polynomial operators. A linear operator in a function space with range contained in Π_m for some m is called a *polynomial operator*. The space of all real functions of bounded variation on I which are left continuous is denoted by $BV(I)$ and it is equipped with the norm

$$\|F\|_{BV(I)} = |F(0)| + \mathrm{var}(F).$$

Moreover, define

$$D(I) = \{F \in BV(I): F \text{ is nondecreasing on } I,\ F(0) = 0,\ F(1) = 1\}.$$

The polynomial operator T_m is now defined for $F \in BV(I)$ by the formula

$$T_m F(x) = \sum_{i=0}^{m} \int_I M_{i,m} dF \int_0^x N_{i,m}(y) dy. \tag{2.1}$$

It then follows that

$$T_m: BV(I) \to \Pi_{m+1} \tag{2.2}$$

and

$$T_m: D(I) \to \Pi_{m+1} \cap D(I). \tag{2.3}$$

The polynomial operators corresponding to the densities are going to be defined naturally by means of the kernel

$$R_m(x, y) = \sum_{i=0}^{m} M_{i,m}(x) N_{i,m}(y). \tag{2.4}$$

It follows by the definitions and properties of $M_{i,m}$ and $N_{i,m}$ that

$$R_m(x, y) = R_m(y, x), \quad 0 \leqslant R_m(x, y) \leqslant m+1 \quad \text{for } x, y \in I. \tag{2.5}$$

Define

$$R_m f(x) = \int_I R_m(x, y) f(y)\, dy.$$

Clearly, $R_m\colon L^2 \to \Pi_m$ and, since by (2.4),

$$R_m f = \sum_{i=0}^{m} (M_{i,m}, f) N_{i,n}, \tag{2.6}$$

it takes by 2° and 3° densities into densities.

It is worth to notice that for F being absolutely continuous (2.1) gives

$$DT_m F = R_m DF. \tag{2.7}$$

PROPOSITION 2.8. *For F in $BV(I)$ we have*

$$T_m F(x) = F(0)(1-x)^{m+1} + F(1)x^{m+1} + \sum_{i=1}^{m} (F, M_{i-1,m-1}) N_{i,m+1}(x).$$

Proof. Direct computation gives

$$\int_I M_{i,m}\, dF = (m+1)\{\delta_{i,m} F(1) - \delta_{i,0} F(0) + (F, M_{i,m-1}) - (F, M_{i-1,m-1})\},$$

and therefore, by 5°,

$$\begin{aligned} T_m F(x) &= F(0) + \sum_{i=0}^{m} \int_I M_{i,m}\, dF \int_0^x N_{i,m} \\ &= F(0)(1-x)^{m+1} + F(1)x^{m+1} + \\ &\quad + \sum_{i=0}^{m-1} (F, M_{i,m-1}) \int_0^x (m+1)\big(N_{i,m}(y) - N_{i+1,m}(y)\big)\, dy \\ &= F(0)(1-x)^{m+1} + F(1)x^{m+1} + \sum_{i=0}^{m-1} (F, M_{i,m-1}) \int_0^x DN_{i+1,m+1}(y)\, dy. \end{aligned}$$

In the next section still a different representation of R_m will be needed. Denote by l_j ($j = 0, \ldots, n$) the orthonormal Legendre polynomials on I, i.e.

$l_j \in \Pi_j$ and $(l_i, l_j) = \delta_{i,j}$ for $i, j = 0, 1, \ldots$, and let

$$\lambda_{o,m} = 1,\ \lambda_{i,m} = \prod_{j=1}^{i} \frac{m+1-j}{m+1+j}, \qquad i = 1, \ldots, m.$$

We do assume in what follows that $\lambda_{i,m} = 0$ for $i = m+1, \ldots$ It is worth to notice that $(\lambda_{i,m})_{i=0,\ldots,m}$ are eigenvalues for the Gram matrix $[(M_{i,m}, N_{j,m})]_{i,j=0,\ldots,m}$. This fact and the representation

$$R_m(x, y) = \sum_{i=0}^{m} \lambda_{i,m} l_i(x) l_i(y) \tag{2.9}$$

are established in [2].

LEMMA 2.10. *Let* $\lambda_n = n(n+1)$ *for* $n = 0, 1, \ldots$ *Then the eigenvalues* $(\lambda_{i,m})_{i=0,\ldots,m}$ *have the following properties*:

(a) $$1 = \lambda_{0,m} > \lambda_{1,m} > \ldots > \lambda_{m,m} > 0;$$

(b) $$\lambda_{i,m+1} \geqslant \lambda_{i,m} \quad \textit{for } i = 1, 2, \ldots;$$

(c) $$\lambda_{i-1,m} - \lambda_{i,m} = \lambda_{i-1,m} \frac{2i}{m+1+i} \quad \textit{for } i = 1, 2, \ldots;$$

(d) $$\frac{1}{2}\frac{\lambda_j}{\lambda_n} < 1 - \lambda_{j,\lambda_n} < \frac{\lambda_j}{\lambda_n} \quad \textit{for } j = 1, \ldots, n.$$

Proof. The computation giving (a)-(c) will be omitted. For the proof of (d) set $m = \lambda_n$ and write

$$\lambda_{j,m} = \prod_{i=1}^{j} (1-a_i), \qquad \text{where } a_i = \frac{2i}{m+1+j}. \tag{2.11}$$

It then follows that $0 < a_i < 1$ for $j = 1, \ldots$ and that

$$A_j := \sum_{i=1}^{j} a_j \leqslant \frac{\lambda_j}{\lambda_n} \leqslant 1. \tag{2.12}$$

Now, (2.11), (2.12) and the Weierstrass inequalities ([4], p. 207) for $j = 1, \ldots, n$ give

$$\lambda_j/\lambda_n > A_j > 1 - \lambda_{j,m},$$

$$1 - \lambda_{j,m} = 1 - \prod_{i=1}^{j} (1-a_i) > 1 - \frac{1}{1+A_j} \geqslant \frac{1}{2} A_j,$$

$$A_j \geqslant \frac{2}{m+1+j} \sum_{i=1}^{j} i = \frac{\lambda_j}{\lambda_n + 1 + j} \geqslant \frac{\lambda_j}{\lambda_n + n + 1} = \frac{n}{n+1} \frac{\lambda_j}{\lambda_n} \geqslant \frac{1}{2} \frac{\lambda_j}{\lambda_n},$$

which completes the proof.

The spectral representation (2.9) gives the formulas

(2.13) $$\|R_m f\|_2^2 = \sum_{i=0}^{m} \lambda_{i,m}^2 a_i^2, \quad \|f-R_m f\|_2^2 = \sum_{i=0}^{\infty} (1-\lambda_{i,m})^2 a_i^2,$$

where $a_i = (f, l_i)$. This and Lemma 2.10 imply

(2.14) $$\|R_m\|_2 \leqslant \|R_{m+1}\|_2 \quad \text{and} \quad \|f-R_{m+1} f\|_2 \leqslant \|f-R_m f\|_2.$$

3. Approximation properties of the polynomial operators. In this section we state the necessary results on approximation by the operators T_m and R_m. The following is a consequence of Proposition 2.8:

COROLLARY 3.1. *For $m = 0, 1, \ldots$ and $F \in BV(I)$ we have*

(3.2) $$\|T_m F\|_\infty \leqslant 3\|F\|_\infty,$$

and, for $F, G \in D(I)$,

(3.3) $$\|T_m F - T_m G\|_\infty \leqslant \|F-G\|_\infty.$$

PROPOSITION 3.4. *For $f \in L^2(I)$ we have*

(3.5) $$\|R_m f\|_2 \leqslant \|f\|_2, \quad m = 0, 1, \ldots,$$

and if $f \in L_2(I)$, then

(3.6) $$\|f-R_m f\|_2 \to 0 \quad \text{as } m \to \infty.$$

For the proof we refer to [1].

PROPOSITION 3.7. *Let $F \in C(I)$. Then $\|F-T_m F\|_\infty \to 0$ as $m \to \infty$.*

Proof. Since (3.2) takes place, it is sufficient to check the statement for absolutely continuous F. However, in this case (2.7) implies, for $f = DF$,

$$|F(x) - T_m F(x)| \leqslant \|DF - DT_m F\|_1 = \|f-R_m f\|_1,$$

and the last term by (3.6) tends to 0 as $m \to \infty$.

For our purpose a characterization of certain Lipschitz classes in terms of the best approximation by Π_n in the L^2-norm will be needed. For given $f \in L^2$ the best approximation is defined by the formula

(3.8) $$E_n(f) = \inf\{\|f-w\|_2 \colon w \in \Pi_n\}.$$

It is well-known that the extreme w in (3.8) is the orthogonal projection of f onto Π_n, i.e.

(3.9) $$E_n^2 = \|f - \sum_{i=0}^{n} (f, l_i) l_i\|_2^2 = \sum_{i=n+1}^{\infty} (f, l_i)^2.$$

In order to define the proper Lipschitz classes following [3], we need the step-weight function

$$\varphi(x) = \sqrt{x(1-x)}, \quad x \in I,$$

and the symmetric difference of the second order:

$$\Delta_h^2 f(x) = f(x+h) - 2f(x) + f(x-h).$$

Now, the modulus of smoothness with the step-weight φ is given by

$$\omega_{2,\varphi}(f;\ \delta) = \sup_{0<h<\delta} \|\Delta_{h\varphi(x)}^2 f(x)\|_2,$$

where $\Delta_{h\varphi(x)}^2$ is zero whenever either $x+h\varphi(x)$ or $x-h\varphi(x)$ is not in I. We are going to use a particular case of Corollary 7.2.1 from [3], i.e.

PROPOSITION 3.10. *Let $f \in L_2(I)$, $0 < \alpha < 2$. Then*

$$\omega_{2,\varphi}(f;\ \delta) = O(\delta^\alpha) \text{ as } \delta \to 0_+ \Leftrightarrow E_n(f) = O\left(\frac{1}{n^\alpha}\right) \text{ as } n \to \infty.$$

To relate this characterization of Lipschitz classes to the approximation by R_m we prove

LEMMA 3.11. *Let $\lambda_n = n(n+1)$ and let $f \in L^2(I)$. Then*

$$\frac{1}{2}E_n(f) \leqslant \|f - R_{\lambda_n} f\|_2 \leqslant E_n(f) + 2\left(\frac{1}{\lambda_n^2}\sum_{i=1}^{n} \lambda_i i E_{i-1}^2(f)\right)^{1/2}. \tag{3.12}$$

Proof. The left-hand side of (3.12) follows by Lemma 2.10(d) and by (2.13). To obtain the right-hand side we use the Abel's transformation

$$\sum_{i=1}^{n} \alpha_i \beta_i = \sum_{i=1}^{n} \left(\sum_{k=i}^{n} \alpha_k\right)(\beta_i - \beta_{i-1}) \quad \text{with } \beta_0 = 0.$$

Defining $a_i = (f, l_i)$ and letting $m = \lambda_n$, we get by (2.13)

$$\|f - R_m f\|_2^2 \leqslant E_n^2(f) + \sum_{i=1}^{n} \alpha_i \beta_i,$$

where $\alpha_i = a_i^2$ and $\beta_i = (1-\lambda_{i,m})^2$. Thus, according to (3.9),

$$\sum_{i=1}^{n} \alpha_i \beta_i \leqslant \sum_{i=1}^{n} E_{i-1}^2 \left[(1-\lambda_{i,m})^2 - (1-\lambda_{i-1,m})^2\right].$$

Now, Lemma 2.10 gives

$$0 \leqslant (1-\lambda_{i,m})^2 - (1-\lambda_{i-1,m})^2 \leqslant (\lambda_{i-1,m} - \lambda_{i,m})(1-\lambda_{i,m} + 1 - \lambda_{i-1,m})$$
$$\leqslant 2\frac{\lambda_i}{\lambda_n}\lambda_{i-1,m}\frac{2i}{m+1+i} \leqslant 4\frac{\lambda_i i}{\lambda_n^2}$$

and this completes the proof.

Now the monotonicity (2.14) and Lemma 3.11 give

COROLLARY 3.13. *Let $f \in L^2(I)$ and let $0 < \alpha < 2$. Then*

$$\|f - R_n f\|_2 = O\left(\frac{1}{n^{\alpha/2}}\right) \text{ as } n \to \infty \Leftrightarrow E_n(f) = O\left(\frac{1}{n^{\alpha}}\right) \text{ as } n \to \infty.$$

Finally, Corollary 3.13 and Proposition 3.10 give

COROLLARY 3.14. *Let α and f be given such that $0 < \alpha < 1$, $f \in L^2(I)$. Then*

$$\|f - R_m f\|_2 = O\left(\frac{1}{m^{\alpha}}\right) \text{ as } m \to \infty \Leftrightarrow \omega_{2,\varphi}(f;\delta) = O(\delta^{2\alpha}) \text{ as } \delta \to 0_+.$$

The author was kindly informed by K. G. Ivanov that Corollary 3.14 has its L^p version which can be proved by methods developed by Berens–Lorentz and Totik. Since the L_p case is more difficult and we need it only for $p = 2$, we restrict our attention here to the L^2 case only.

4. The estimators. Let us start with a simple sample of size n: $X_1, \ldots, X_n$. It is assumed that the common distribution function F of these i.i.d. random variables has its support in I. For the given sample let us introduce

$$f_{m,n}(x) = \frac{1}{n} \sum_{j=0}^{n} R_m(X_j, x), \quad x \in I. \tag{4.1}$$

$f_{m,n}$ is a polynomial of degree m which, by (2.6), is a density on I. Let now F_n be the empirical distribution, i.e. $F_n = |\{i\colon X_i < x\}|/n$, and let

$$F_{m,n} = T_m F_n. \tag{4.2}$$

It follows by (2.1) that

$$DF_{m,n} = f_{m,n}. \tag{4.3}$$

PROPOSITION 4.4. *Let F and $X_1, X_2, \ldots$ be given as above. Then $\mathrm{P}\{F_{m,n} \Rightarrow F$ as $m, n \to \infty\} = 1$, where $\Rightarrow$ means the weak convergence of probability distribution functions.*

Proof. Let us start with the following identity:

$$F - F_{m,n} = (F - T_m F) + T_m(F - F_n). \tag{4.5}$$

It will be shown at first that $T_m F$ converges weakly to F as $m \to \infty$ for each $F \in D(I)$. For φ continuous on $(-\infty, \infty)$ and with compact support, according to (2.1) and (3.6) we obtain

$$\int_{-\infty}^{\infty} \varphi\, dT_m F = \int_0^1 R_m(\varphi|_I)\, dF \to \int_{-\infty}^{\infty} \varphi\, dF \quad \text{as } m \to \infty.$$

For the second part of (4.5) we obtain by (3.3) that

$$\|T_m(F-F_n)\|_\infty \leqslant \|F-F_n\|_\infty,$$

but, by Glivenko's theorem (see [5]),

$$\mathrm{P}\{\|F-F_n\|_\infty \to 0 \text{ as } n\to\infty\} = 1.$$

Thus, with probability 1, $T_m(F-F_n)$ tends uniformly on I to 0 as m, $n\to\infty$. Since $F(x)-T_m F(x)\to 0$ as $m\to\infty$ at each continuity point of F, it follows by (4.5) that, with probability 1, $F_{m,n}(x)\to F(x)$ at each such a point.

PROPOSITION 4.6. *Let $F\in D(I)\cap C(I)$. Then*

$$\mathrm{P}\{\|F-F_{m,n}\|_\infty \to 0 \text{ as } m,\ n\to\infty\} = 1.$$

This follows from the proof of Proposition 4.4 and by Proposition 3.7.

We now pass to the density estimation. In what follows we do assume that the density $f = DF$ is in $L^2(I)$.

LEMMA 4.7. *For the density $f\in L^2$ we have*

$$\|f-R_m f\|_2^2 \leqslant E\|f-f_{m,n}\|_2^2 \leqslant \|f-R_m f\|_2^2+\frac{m+1}{n}.$$

Proof. Note that $Ef_{m,n} = R_m f(x)$ and, therefore, by Jensen's inequality the left-hand side follows. On the other hand,

$$E\|f-f_{m,n}\|_2^2 = \|f\|_2^2-2(f,\ R_m f)+E\|f_{m,n}\|_2^2$$

$$= \|f-R_m f\|_2^2+\frac{1}{n}\iint_{I^2} R_m^2(y,\ x)\,f(y)\,dx\,dy-\frac{1}{n}\|R_m f\|_2^2.$$

Since, by (2.4), $R_m(x,\ y)\leqslant m+1$, the desired inequality follows.

MAIN THEOREM. *Let the density be in $L^2(I)$. Let α, $0<\alpha<1$, be given. Define $f_n = f_{m,n}$, where $m = [n^\beta]$, $\beta = 1/(2\alpha+1)$, and $[\ ]$ denotes the integer part. Then the following conditions are equivalent:*

(i) $$\omega_{2,\varphi}(f;\ \delta) = O(\delta^{2\alpha}) \quad \text{as } \delta\to 0_+,$$

(ii) $$E\|f-f_n\|_2^2 = O\left(\frac{1}{n^{2\alpha/(1+2\alpha)}}\right) \quad \text{as } n\to\infty.$$

Proof. According to Lemma 4.7

$$E\|f-f_n\|_2^2 \leqslant \|f-R_m f\|_2^2+\frac{m+1}{n}, \quad \text{where } m = [n^\beta].$$

Applying Corollary 3.14 we get the implication (i)$\Rightarrow$(ii). Conversely, from Lemma 4.7 we deduce that (ii) implies

$$\|f-R_m f\|_2^2 = O\left(\frac{1}{n^{2\alpha/(1+2\alpha)}}\right) \quad \text{for } m = [n^\beta],$$

whence

$$\|f-R_m f\|_2^2 = O\left(\frac{1}{m^\alpha}\right) \quad \text{as } m \to \infty.$$

This and Corollary 3.14 complete the proof.

5. Algorithm for computing the density and distribution estimators. Let $X_1, \ldots, X_n$ be given as in the previous section. Since

$$f_{m,n} = \frac{1}{n}\sum_{j=1}^{n} R_m(X_j, x),$$

to compute $f_{n,m}(x)$ for fixed x we need to compute $R_m(X_j, x)$ for $j = 1, \ldots, n$. However,

$$R_m(X_j, x) = \sum_{i=0}^{m} M_{i,m}(X_j) N_{i,m}(x)$$

and, therefore, we use the Casteljeau algorithm for the first time to compute $M_{i,m}(X_j)$ and for the second time to calculate $R_m(X_j, x)$. Now, for the density $f_{m,n}$ we have also the representation

$$f_{m,n}(x) = \sum_{i=0}^{m} a_i N_{i,m}(x),$$

where

$$a_i = \frac{1}{n}\sum_{j=1}^{n} M_{i,m}(X_j), \quad i = 0, \ldots, m.$$

Thus, at almost no cost the coefficients

$$b_0 = 0,\ b_1 = 1,\ b_j = \frac{a_0 + \ldots + a_{j-1}}{m+1}, \quad j = 1, \ldots, m+1,$$

can be computed. To compute $F_{m,n}(x)$ one applies once more the Casteljeau algorithm to the formula

$$F_{m,n}(x) = \int_0^x f_{m,n}(y)\,dy = \sum_{j=1}^{m+1} b_j M_{j,m+1}(x).$$

Added in proof. The author has recently learned that (2.9), p. 4, was established earlier by M. M. Derriennic in J. of Appr. Theory 31 (1981), p. 337.

REFERENCES

[1] Z. Ciesielski, *Approximation by polynomials and extension of Parseval's identity for Legendre polynomials to the L_p case*, Acta Scient. Math. 48 (1985), p. 65-70.
[2] Z. Ciesielski and J. Domsta, *The degenerate B-splines as basis in the space of algebraic polynomials*, Ann. Polon. Math. 46 (1985), p. 71-79.
[3] Z. Ditzian and V. Totik, *Moduli of smoothness*, Dept. of Math., Univ. of Alberta, Edmonton (1986), 477 p., preprint.
[4] D. S. Mitrinovic, *Analiticke nejednakosti*, Beograd 1970.
[5] A. Rényi, *Wahrscheinlichkeitsrechnung*, Berlin 1962.

Instytut Matematyczny PAN
ul. Abrahama 18
81-825 Sopot, Poland

Received on 15. 12. 1986