Journal of Statistical Planning and Inference 105 (2002) 377–392

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Application of Bernstein Polynomials for smooth estimation of a distribution and density function

G. Jogesh Babu[a], Angelo J. Canty[b], Yogendra P. Chaubey[b,*]

[a]*Department of Statistics, Pennsylvania State University, State College, PA, USA*

[b]*Department of Mathematics and Statistics, Concordia University, Loyola Campus, 7141 Sherbrooke St. W., Montreal, Que., Canada H4B 1R6*

Received 14 September 2000; accepted 20 July 2001

Abstract

The empirical distribution function is known to have optimum properties as an estimator of the underlying distribution function. However, it may not be appropriate for estimating continuous distributions because of its jump discontinuities. In this paper, we consider the application of Bernstein polynomials for approximating a bounded and continuous function and show that it can be naturally adapted for smooth estimation of a distribution function concentrated on the interval $[0, 1]$ by a continuous approximation of the empirical distribution function. The smoothness of the approximating polynomial is further used in deriving a smooth estimator of the corresponding density. The asymptotic properties of the resulting estimators are investigated. Specifically, we obtain strong consistency and asymptotic normality under appropriate choice of the degree of the polynomial. The case of distributions with other compact and non-compact support can be dealt through transformations. Thus, this paper gives a general method for non-parametric density estimation as an alternative to the current estimators. A small numerical investigation shows that the estimator proposed here may be preferable to the popular kernel-density estimator.

MSC: 62G05; 62G07

Keywords: Kernel estimator; Non-parametric density estimator; Strong consistency

1. Introduction

There has been a considerable development of methods for smooth estimation of density and distribution functions, following the introduction of kernel method of smoothing by Rosenblatt (1956) and the further advances made on kernel method by Parzen (1962). The reader is referred to the excellent texts by Härdle (1991) and Silverman (1986) for smoothing techniques for curve estimation. These methods, however, do not

* Corresponding author. Fax: +1-514-848-2831.
E-mail address: chaubey@alcor.concordia.ca (Y.P. Chaubey).

PII: S0378-3758(01)00265-8

take advantage of the knowledge of the support of the distribution and work as if the support is the whole real line. Asymptotically it does not matter, but in specific cases one may see a positive probability associated with zero probability region using the kernel methods.

Chaubey and Sen (1996) considered smooth estimation of the survival function $S(x)=1-F(x)$ for life time random variable X based on a random sample $X_1, X_2, \ldots, X_n$, as an alternative to Nadaraya–Watson estimator. Here F denotes the distribution function of the random variable X, which typically has support $\mathbf{R}^+$. This estimator is based on smoothing the empirical distribution function, F_n given by

$$F_n(x)=n^{-1}\sum_{i=1}^{n} I\{X_i \leqslant x\} \tag{1.1}$$

and is appropriate when the support of the distribution is $\mathbf{R}^+$. However, when the support of F is contained in compact interval $[a,b]$, $a<b$, it is simpler to transform the random variable X to Y with support $[0,1]$ through the transformation $Y=(X-a)/(b-a)$. Subsequently, we can adapt the following theorem using Bernstein polynomials for smoothing the empirical distribution function over the interval $[0,1]$. Transformations such as $Y=X/(1+X)$ and $Y=(\frac{1}{2})+(1/\pi)\tan^{-1}X$ can be used also to cover the cases of random variables with support $[0,\infty)$ and $(-\infty,\infty)$, respectively.

Theorem 1.1 (Feller, 1965, Theorem 1, Section VII.2). *If $u(x)$ is a bounded and continuous function on the interval $[0,1]$, then as $m \to \infty$*

$$u_m^*(x)=\sum_{k=0}^{m} u(k/m)b_k(m,x) \to u(x) \tag{1.2}$$

uniformly for $x \in [0,1]$, where

$$b_k(m,x)=\binom{m}{k} x^k(1-x)^{m-k}, \quad k=0,\ldots,m. \tag{1.3}$$

In this paper, we consider only distributions F with support $[0,1]$. For such functions Theorem 1.1 motivates the smooth estimator,

$$\tilde{F}_{n,m}(x)=\sum_{k=0}^{m} F_n(k/m)b_k(m,x), \quad x\in[0,1], \tag{1.4}$$

based on Bernstein polynomials. Note that $\tilde{F}_{n,m}$ is a polynomial in x, and hence has all the derivatives.

We now demonstrate that $\tilde{F}_{n,m}$ is a proper distribution function, and hence it qualifies as an estimator of F. Clearly $\tilde{F}_{n,m}$ is continuous, $0 \leqslant \tilde{F}_{n,m}(x) \leqslant 1$ for $x\in[0,1]$, and

$$\tilde{F}_{n,m}(x)=\sum_{k=0}^{m} f_n(k/m)B_k(m,x), \tag{1.5}$$

where

$$f_n(0)=0, \quad f_n(k/m)=F_n(k/m)-F_n((k-1)/m), \quad k=1,\ldots,m \tag{1.6}$$

and

$$B_k(m,x) = \sum_{j=k}^{m} b_k(m,x) = m \binom{m-1}{k-1} \int_0^x t^{k-1}(1-t)^{m-k}\,\mathrm{d}t. \tag{1.7}$$

For all k, $f_n(k/m)$ is non-negative by (1.6), and $B_k(m,x)$ is non-decreasing in x by (1.7). Thus $\tilde{F}_{n,m}$ is non-decreasing in x.

Furthermore, representation (1.5) leads to a smooth estimator

$$\begin{aligned}\tilde{f}_{n,m}(x) &= \sum_{k=1}^{m} f_n(k/m)\frac{\mathrm{d}}{\mathrm{d}x}B_k(m,x)\\ &= m\sum_{k=0}^{m-1} f_n((k+1)/m)b_k(m-1,x)\\ &= m\sum_{k=0}^{m-1} (F_n((k+1)/m) - F_n(k/m))b_k(m-1,x) \end{aligned}\tag{1.8}$$

for density f of F. The asymptotic properties of $\tilde{F}_{n,m}$ and $\tilde{f}_{n,m}$ are examined in the next two sections.

2. Asymptotic properties of $\tilde{F}_{n,m}$

Throughout this paper we use the notation

$$||G|| = \sup_{x\in[0,1]} |G(x)|$$

for a bounded function G on $[0,1]$,

$$a_n = (n^{-1}\log n)^{1/2} \quad \text{and} \quad b_{n,m} = (n^{-1}\log n)^{1/2}(m^{-1}\log m)^{1/4}. \tag{2.1}$$

The following theorem shows that $\tilde{F}_{n,m}$ is strongly consistent.

Theorem 2.1. *Let F be a continuous probability distribution function on the interval $[0,1]$. If $m,n\to\infty$, then $||\tilde{F}_{n,m} - F|| \to 0$ a.s.*

Proof. Recall definition (1.2) of u_m^* for any u. We first note that for all $x\in[0,1]$,

$$||\tilde{F}_{n,m} - F|| \leqslant ||\tilde{F}_{n,m} - F_m^*|| + ||F_m^* - F||. \tag{2.2}$$

As

$$\tilde{F}_{n,m}(x) - F_m^*(x) = \sum_{k=0}^{m} (F_n(k/m) - F(k/m))b_k(m,x),$$

we have

$$||\tilde{F}_{n,m} - F_m^*|| \leqslant \max_{0\leqslant k\leqslant m} |F_n(k/m) - F(k/m)| \leqslant ||F_n - F||. \tag{2.3}$$

Since by Glivenko–Cantelli theorem, $||F_n - F|| \to 0$ a.s. as $n\to\infty$, the result follows from (2.2), (2.3) and Theorem 1.1. $\square$

We shall now examine the closeness of the smooth estimator with the empirical distribution function, as it has many optimal properties. Recall that a function g is said to be Lipschitz of order α, if there exists a constant K such that

$$|g(s) - g(t)| \leqslant K|s-t|^{\alpha}.$$

Theorem 2.2. *Let F be continuous and differentiable on the interval $[0,1]$ with density f. If f is Lipschitz of order 1, then for $n^{2/3} \leqslant m \leqslant (n/\log n)^2$, we have a.s. as $n \to \infty$,*

$$||\tilde{F}_{n,m} - F_n|| = \mathrm{O}((n^{-1}\log n)^{1/2}(m^{-1}\log m)^{1/4}). \tag{2.4}$$

Thus for $m = n$, we have

$$||\tilde{F}_{n,n} - F_n|| = \mathrm{O}(n^{-3/4}(\log n)^{3/4}) \quad \text{a.s.} \tag{2.5}$$

To prove the theorem we need the following lemmas.

Lemma 2.1 (Babu, 1989). *Let $Z_1,\ldots,Z_n$ be independent random variables with mean zero and let $|Z_i| \leqslant b$ for some $b > 0$. If $V \geqslant \sum_{i=1}^n E(Z_i^2)$, then for all $0 < s < 1$ and $0 \leqslant a \leqslant V/sb$, we have*

$$P\left(\sum_{i=1}^n Z_i > a\right) \leqslant \exp(-a^2 s(1-s)/V). \tag{2.6}$$

Lemma 2.2. *Let F be Lipschitz of order 1 and define*

$$N_{x,m} = \{0 \leqslant k \leqslant m : |k - xm| \leqslant (m\log m)^{1/2}\}. \tag{2.7}$$

Then for $2 \leqslant m \leqslant (n/\log n)^2$, we have $H_{n,m} = \mathrm{O}(b_{n,m})$ a.s. as $n \to \infty$, where

$$H_{n,m} = \sup_{0<x<1}\max_{k\in N_{x,m}} |F_n(k/m) - F(k/m) - F_n(x) + F(x)|. \tag{2.8}$$

Proof. By dividing the unit interval into sub-intervals of length $b_{n,m}$ and noting that $b_{n,m} \leqslant a_m$ for $m \leqslant (n/\log n)^2$, we have

$$H_{n,m} \leqslant D_{n,m} + D_{n,m,1}, \tag{2.9}$$

where

$$D_{n,m} = \max_{|i-j|b_{n,m} \leqslant 2a_m} |F_n(jb_{n,m}) - F_n(ib_{n,m}) - F(jb_{n,m}) + F(ib_{n,m})|$$

and

$$D_{n,m,1} = \max_{i,j} |F((j+1)b_{n,m}) - F(jb_{n,m}) + F((i+1)b_{n,m}) - F(ib_{n,m})|.$$

Since F is Lipschitz of order 1, we have for $0 \leqslant u,\ v \leqslant 1$,

$$|F(u) - F(v)| \leqslant c|u-v| \quad \text{for some } c \geqslant 2 \tag{2.10}$$

and hence

$$D_{n,m,1} = \mathrm{O}(b_{n,m}). \tag{2.11}$$

We shall now show that $D_{n,m}=\mathrm{O}(b_{n,m})$ a.s. as $n\to\infty$. For $u\leqslant v$, let

$$Z_{i,u,v}=I_{u<X_i\leqslant v}-(F(v)-F(u)).$$

By (2.10), $\sum_{i=1}^n \mathrm{Var}(Z_{i,u,v})=n\,\mathrm{Var}(Z_{1,u,v})\leqslant cn|u-v|$. For $|u-v|\leqslant 2a_m$, we apply Lemma 2.1 with $Z_i=Z_{i,u,v}$, $b=1$, $s=1/2$, $a=4cnb_{n,m}$, and $V=2cna_m$, to get

$$\begin{aligned}P(|F_n(u)-F_n(v)-F(u)+F(v)|>4cb_{n,m}) &\leqslant 2\exp(-a^2/4V)\\ &\leqslant 2\exp(-2cnb_{n,m}^2/a_m)\\ &\leqslant 2\exp(-2c\log n)\\ &=2n^{-2c}.\end{aligned}\tag{2.12}$$

As $b_{n,m}^{-1}\leqslant n$ and # of pairs $\{(i,j)\colon 0<ib_{n,m}\leqslant 1, 0<jb_{n,m}\leqslant 1\}\leqslant b_{n,m}^{-2}$, it follows by (2.12) that

$$P(D_{n,m}>4cb_{n,m})\leqslant 2n^2n^{-2c}\leqslant 2n^{-2}.$$

Therefore,

$$\sum_n P(D_{n,m}>4cb_{n,m})<\infty.$$

Hence by (2.9), (2.11) and Borel–Cantelli lemma, we have $H_{n,m}=\mathrm{O}(b_{n,m})$ a.s. as $n\to\infty$. □

Proof of Theorem 2.2. Observe that for every $x\in[0,1]$,

$$\begin{aligned}\tilde F_{n,m}(x)-F_n(x)=&\sum_{k=0}^m b_k(m,x)(F_n(k/m)-F(k/m)-F_n(x)+F(x))\\ &+(F_m^*(x)-F(x)).\end{aligned}\tag{2.13}$$

As f is Lipschitz of order 1, we have uniformly in $0<x<1$, that

$$F(k/m)-F(x)=((k/m)-x)f(x)+\mathrm{O}(((k/m)-x)^2).\tag{2.14}$$

Since,

$$\sum_{k=0}^m kb_k(m,x)=mx \quad\text{and}\quad \sum_{k=0}^m (k-mx)^2b_k(m,x)=mx(1-x),\tag{2.15}$$

it follows by (2.14) that

$$||F_m^*-F||=\mathrm{O}(1/m).\tag{2.16}$$

By applying Lemma 2.1 with Z_i i.i.d. random variables satisfying $P(Z_i=1-x)=x=1-P(Z_i=-x)$, $a=(m\log m)^{1/2}$, $V=m/4$ and $s=\frac{1}{2}$, we get for $m\geqslant 16$,

$$\sum_{k=0,k\notin N_{x,m}}^m b_k(m,x)\leqslant\frac{2}{m}.\tag{2.17}$$

The theorem now follows from (2.13), (2.16), (2.17) and Lemma 2.2 by noting that $m^{-1} \leqslant b_{n,m}$ for $n^{2/3} \leqslant m$. $\square$

Remark 2.1. Results similar to Theorem 2.2 for mixing sequences of random variables can be obtained using Lemmas 2.1 and 3.3 of Babu and Singh (1978).

3. Asymptotic properties of $\tilde{f}_{n,m}$

We now establish a strong convergence result for $\tilde{f}_{n,m}$ similar to that of $\tilde{F}_{n,m}$. Throughout this section, we assume that F has continuous density f, and that f is Lipschitz of order 1. Let

$$M = m - 1 \quad \text{and} \quad \gamma(x) = f(x)(4\pi x(1-x))^{-1/2} \text{ for } 0 < x < 1. \tag{3.1}$$

Theorem 3.1. *For $2 \leqslant m \leqslant (n/\log n)$, we have a.s. as $n \to \infty$,*

$$||\tilde{f}_{n,m} - f|| = \mathrm{O}(m^{1/2} a_n) + \mathrm{O}(||F_m^{*'} - f||), \tag{3.2}$$

where $F_m^{*'}$ denotes the derivative of F_m^*. Consequently, if $m = \mathrm{o}(n/\log n)$, then $||\tilde{f}_{n,m} - f|| \to 0$ a.s. as $m, n \to \infty$.

Proof. First, we write

$$\begin{aligned}
\tilde{f}_{n,m}(x) &= m \sum_{k=0}^{M} (F_n((k+1)/m) - F_n(k/m)) b_k(M; x) \\
&= m \sum_{k=0}^{M} (F((k+1)/m) - F(k/m)) b_k(M; x) \\
&\quad + m \sum_{k=0}^{M} (F_n((k+1)/m) - F((k+1)/m) \\
&\quad - F_n(k/m) + F(k/m)) b_k(M; x) \\
&= T_m(x) + T_{m,n}(x), \quad \text{say.}
\end{aligned} \tag{3.3}$$

To estimate the last term in (3.3), note that $F((k+1)/m) - F(k/m) \leqslant cm^{-1}$ for some $c > 2$. We now apply Lemma 2.1 with i.i.d. random variables

$$Z_i = I_{(k < mX_i \leqslant k+1)} - (F((k+1)/m) - F(k/m)),$$

$b = 1$, $s = 1/2$, $a = 2cna_n m^{-1/2}$ and $V = (cn/m)$, to obtain

$$P(L_{m,n} > 2ca_n m^{-1/2}) \leqslant 2m\mathrm{e}^{-a^2/4V} \leqslant 2n^{1-c}, \tag{3.4}$$

where

$$L_{m,n} = \sup_{0 \leqslant k \leqslant m-1} |F_n((k+1)/m) - F((k+1)/m) - F_n(k/m) + F(k/m)|.$$

Here we used the condition $m \leqslant (n/\log n)$ to satisfy the requirement $a \leqslant 2V$. Thus $L_{n,m} = \mathrm{O}(a_n m^{-1/2})$ a.s. as $n \to \infty$, which in turn implies

$$||T_{m,n}|| = \mathrm{O}(m^{1/2} a_n)$$

a.s. as $n \to \infty$.

Note that T_m is the derivative of F_m^*. As the density f is assumed to be Lipschitz of order 1, we have uniformly for $0 \leqslant x \leqslant 1$ that

$$\begin{aligned} T_m(x) - f(x) &= \sum_{k=0}^{M} (f(k/m) - f(x)) b_k(M; x) + \mathrm{O}(m^{-1}) \\ &= \mathrm{O}\left(\sum_{k=0}^{M} |(k/m) - x| b_k(M; x) \right) + \mathrm{O}(m^{-1}) \\ &= \mathrm{O}(m^{-1/2}). \end{aligned} \tag{3.5}$$

This completes the proof of Theorem 3.1. □

The following theorems establish asymptotic normality of $\tilde{f}_{n,m}(x)$.

Theorem 3.2. *If $f(x) > 0$, then*

$$n^{1/2} m^{-1/4} (\tilde{f}_{n,m}(x) - f(x)) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \gamma(x)) \tag{3.6}$$

as $m, n \to \infty$ such that $2 \leqslant m \leqslant (n/\log n)$ and $n^{2/3}/m \to 0$.

Theorem 3.3. *Suppose that the density f admits two continuous derivatives f' and f'' in an open neighborhood of x in $[0,1]$. If $f(x) > 0$ and $m, n \to \infty$ such that $\lim_{n\to\infty} n^{-1/5} m^{1/2} \to \delta$ for some $\delta > 0$, then we have*

$$n^{2/5} (\tilde{f}_{n,m}(x) - f(x)) - \tfrac{1}{2}\delta^{-2} b(x) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \gamma(x)\delta), \quad \text{as } n \to \infty, \tag{3.7}$$

where

$$b(x) = (1 - 2x) f'(x) + x(1 - x) f''(x).$$

The non-random part $T_m(x)$ of the decomposition (3.3) of $\tilde{f}_{n,m}(x)$ contributes to the bias. However, the common main part of the proofs involve establishing asymptotic normality of $T_{m,n}(x)$. Since it is of independent interest, the result is presented as a separate proposition. For clarity of presentation we define

$$a_{k,m} = F((k+1)/m) - F(k/m), \tag{3.8}$$

$$Y_{i,m} = \sum_{k=0}^{M} (I_{(k/m < X_i \leqslant (k+1)/m)} - a_{k,m}) b_k(M, x) \tag{3.9}$$

and note that

$$T_{m,n}(x) = \frac{m}{n} \sum_{i=1}^{n} Y_{i,m}. \tag{3.10}$$

Since for each $m, Y_{i,m}$ are i.i.d. random variables, we have

$$\mathrm{Var}(T_{m,n}(x)) = m^2 n^{-1}\,\mathrm{Var}(Y_{1,m}).$$

Proposition 1. *If* $f(x) > 0$, *then*

$$n^{1/2} m^{-1/4} T_{m,n} \xrightarrow{\mathscr{D}} \mathscr{N}(0, \gamma(x)),$$

as $m, n \to \infty$ *such that* $m/n \to 0$.

Proof of Theorem 3.2. By (3.3) and Proposition 1 it is enough to show that

$$T_m(x) - f(x) = \mathrm{o}(n^{-1/2} m^{1/4}).$$

This clearly holds in view of (3.5), as the limit of $n^{2/3}/m$ is assumed to be 0. □

Proof of Theorem 3.3. We first estimate the bias $n^{2/5}(T_m(x) - f(x))$. On $N_{x,m}$ we have

$$\begin{aligned} ma_{k,m} &= f(k/m) + \frac{1}{2m} f'(k/m) + \mathrm{O}(m^{-2}) \\ &= f(x) + \frac{1}{2m} f'(x) + \mathrm{o}(m^{-1}) \end{aligned} \tag{3.11}$$

and

$$\begin{aligned} m(f(k/m) - f(x)) &= (k - xm) f'(x) + \frac{1}{2m}(k - xm)^2 f''(x)(1 + \mathrm{o}(1)) \\ &= (k - Mx) f'(x) - x f'(x) \\ &\quad + \frac{1}{2m}(k - xm)^2 f''(x)(1 + \mathrm{o}(1)). \end{aligned} \tag{3.12}$$

From (2.15), (2.17), (3.11) and (3.12), a simple algebra leads to

$$2m(T_m(x) - f(x)) = b(x) + \mathrm{o}(1). \tag{3.13}$$

The result now follows from (3.13) and Proposition 1, as $n^{-1/5} m^{1/2} \to \delta$. □

Remark 3.1. Clearly, if f has a derivative f' that satisfies Lipschitz condition of order 1, then

$$\begin{aligned} T_m(x) - f(x) &= \sum_{k=0}^{M} (f(k/m) - f(x)) b_k(M; x) + \mathrm{O}(m^{-1}) \\ &= f'(x) \sum_{k=0}^{M} ((k/m) - x) b_k(M; x) \\ &\quad + \mathrm{O}\left(\sum_{k=0}^{M} ((k/m) - x)^2 b_k(M; x) \right) + \mathrm{O}(m^{-1}) \\ &= \mathrm{O}(m^{-1}). \end{aligned}$$

Thus from the proof of Theorem 3.2, if $2 \leqslant m \leqslant (n/\log n)$ and $mn^{-2/5} \to \infty$, then

$$n^{1/2}m^{-1/4}(\tilde{f}_{n,m}(x) - f(x)) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \gamma(x)). \tag{3.14}$$

To prove the proposition we need estimates of $\mathrm{Var}(Y_{1,m})$ and that of the sum of squares of binomial probabilities. These are given in the next two lemmas.

Lemma 3.1. *For any* $0 < x < 1$, *as* $r \to \infty$,

$$2\sqrt{\pi r x(1-x)} \sum_{k=0}^{r} b_k^2(r,x) \to 1.$$

Proof. Let U_i, W_j, $i,j = 1,\dots,r$, be i.i.d. Bernoulli random variables with $P(U_1 = 1) = x = 1 - P(U_1 = 0)$, and let $R_i = (U_i - W_i)/\sqrt{2x(1-x)}$. Then clearly, $E(R_i) = 0$, $\mathrm{Var}(R_i) = 1$, and

$$\sum_{k=0}^{r} b_k^2(r,x) = P\left(\sum_{i=1}^{r} R_i = 0\right).$$

The result now follows from Theorem 3 of Section XV.5 of Feller (1965, p. 490). □

Lemma 3.2. *If* $f(x) > 0$, *then* $m^{3/2}\,\mathrm{Var}(Y_{1,m}) \to \gamma(x)$ *as* $m \to \infty$.

Proof. It is easy to see that

$$\mathrm{Var}(Y_{1,m}) = \sum_{k=0}^{M} b_k^2(M,x)a_{k,m} - \left(\sum_{k=0}^{M} b_k(M,x)a_{k,m}\right)^2. \tag{3.15}$$

Note that by (3.5)

$$m\sum_{k=0}^{M} b_k(M,x)a_{k,m} = T_m(x) = f(x) + \mathrm{O}(m^{-1/2}) = \mathrm{O}(1). \tag{3.16}$$

Furthermore, we have

$$ma_{k,m} = f(\xi) \tag{3.17}$$

for some ξ such that $(k/m) \leqslant \xi \leqslant ((k+1)/m)$. By Lipschitz continuity,

$$\begin{aligned} ma_{k,m} &= f(k/m) + \mathrm{O}(1/m) \\ &= f(x) + \mathrm{O}(|(k/m) - x|) + \mathrm{O}(1/m). \end{aligned} \tag{3.18}$$

Since $b_k(M,x) \leqslant 1$, we have by Cauchy–Schwarz inequality and Lemma 3.1, that

$$\begin{aligned} \left(\sum_{k=0}^{M} b_k^2(M,x)|(k/m) - x|\right)^2 &\leqslant \sum_{k=0}^{M} b_k^3(M,x) \sum_{k=0}^{M} b_k(M,x)((k/m) - x)^2 \\ &= \mathrm{O}\left(\frac{1}{m}\sum_{k=0}^{M} b_k^3(M,x)\right) \\ &= \mathrm{O}(m^{-3/2}). \end{aligned} \tag{3.19}$$

Hence (3.18) and another application of Lemma 3.1 yield,

$$m\sum_{k=0}^{M} b_k^2(M,x)a_{k,m} = f(x)\sum_{k=0}^{M} b_k^2(M,x) + \mathrm{O}(1/m) + \mathrm{O}\left(\sum_{k=0}^{M} b_k^2(M,x)|(k/m)-x|\right) = m^{-1/2}\gamma(x)(1+\mathrm{o}(1)) + \mathrm{O}(m^{-3/4}). \tag{3.20}$$

Now the result follows from (3.15), (3.16) and (3.20). □

Proof of Proposition 1. For the sequence of i.i.d. random variables, we shall first verify Lindeberg condition,

$$s_{n,m}^{-2}\sum_{i=1}^{n} E(Y_{i,m}^2 I_{(|Y_{i,m}|>\varepsilon s_{n,m})}) \to 0 \tag{3.21}$$

for every $\varepsilon > 0$ as $n \to \infty$, where $s_{n,m}^2 = n\,\mathrm{Var}(Y_{1,m})$. By (3.17) and Lemma 3.1, we have

$$|Y_{i,m}| \leqslant \max_{0\leqslant k\leqslant M}(a_{k,m} + b_k(M,x)) = \mathrm{O}(1/m) + \left(\sum_{k=0}^{M} b_k^2(M,x)\right)^{1/2} = \mathrm{O}(m^{-1/4}).$$

As by Lemma 3.2, $s_{n,m}^2 = nm^{-3/2}\gamma(x)(1+\mathrm{o}(1))$, we have

$$|Y_{i,m}|s_{n,m}^{-1} = \mathrm{O}(m^{-1/4}m^{3/4}n^{-1/2}) = \mathrm{O}((m/n)^{1/2}) \to 0,$$

as $m, n \to \infty$ such that $m/n \to 0$. Thus (3.21) holds and hence

$$m^{3/4}n^{-1/2}\sum_{i=1}^{n} Y_{i,m} \xrightarrow{\mathscr{D}} \mathscr{N}(0,\gamma(x)).$$

In view of (3.10), this completes the proof of Proposition 1. □

4. Numerical illustrations

As indicated earlier, the method proposed here may be used for any finite or infinite support using the transformation method. To make the method applicable in general, we list below suggested transformations in different cases.

(1) Suppose X is concentrated on a finite support $[a,b]$, then we work with the sample values $Y_1,\ldots,Y_n$, where $Y_i = (X_i - a)/(b-a)$. Denoting the $\tilde{G}_n(y)$ as the smooth distribution function of the transformed sample, the smooth distribution $\tilde{F}_n(x)$ is

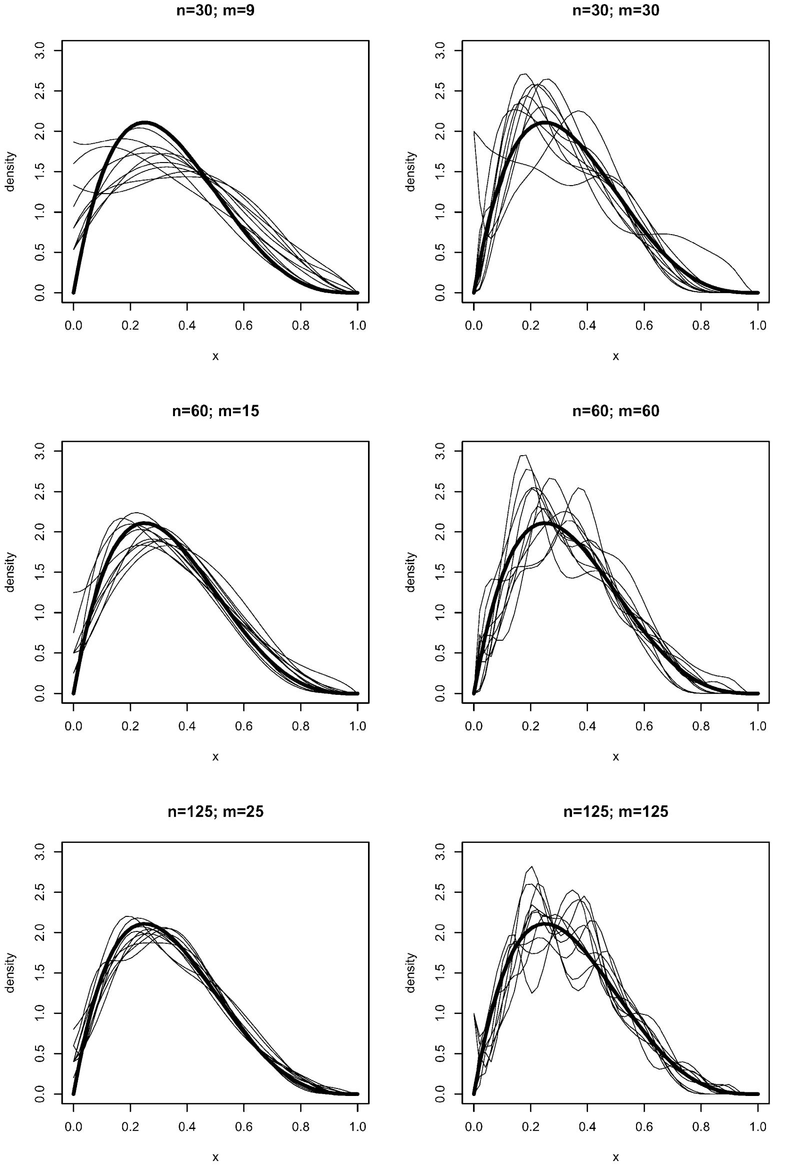

Fig. 1. Smooth Bernstein Density Estimator for Beta(2, 4) Sample.

given by

$$\tilde{F}_n(x) = \tilde{G}_n(y), \quad \text{where } y = (x - a)/(b - a).$$

(2) For the distributions concentrated on the interval $(-\infty, \infty)$, we can use the transformed sample $Y_i = (1/2) + (1/\pi)\tan^{-1} X_i$ which transforms the range to the interval

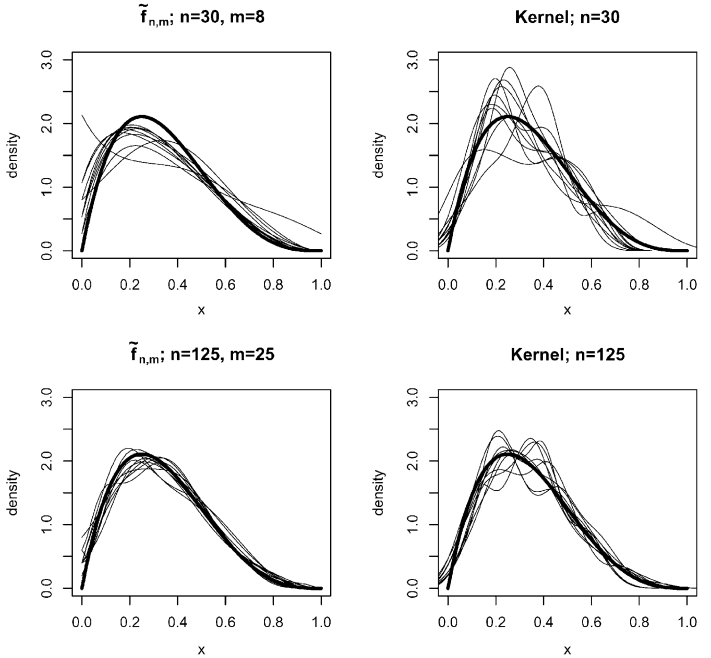

Fig. 2. Bernstein and Kernel Density Estimators for Beta(2, 4) Sample.

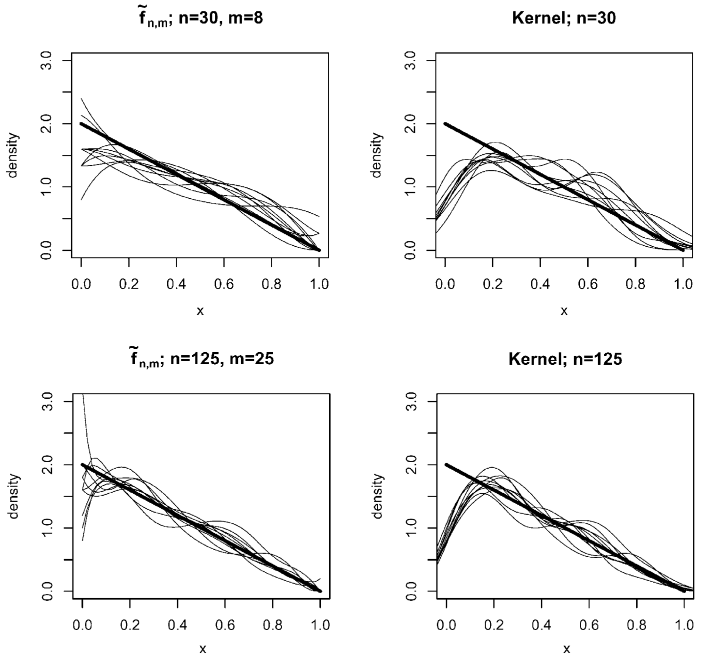

Fig. 3. Bernstein and Kernel Density Estimators for Beta(1, 2) Sample.

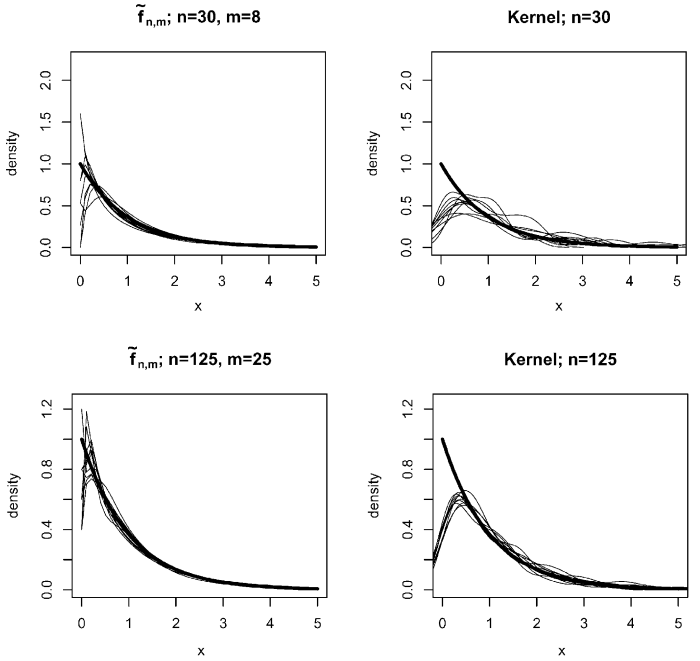

Fig. 4. Bernstein and Kernel Density Estimators for Exp(1) Sample.

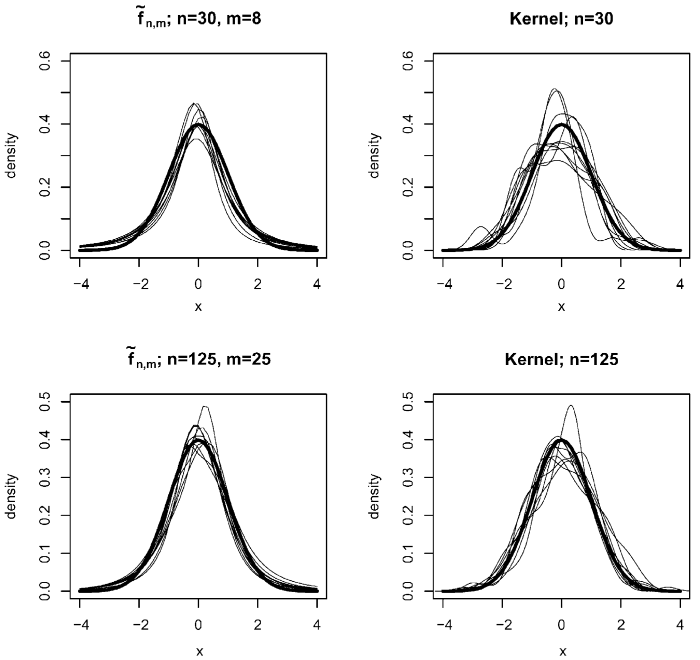

Fig. 5. Bernstein and Kernel Density Estimators for N(0, 1) Sample.

$(0,1)$. Hence

$$\tilde{F}_n(x) = \tilde{G}_n(y), \quad \text{where } y = (1/2) + (1/\pi)\tan^{-1} x.$$

(3) For the non-negative support $[0,\infty)$, the transformation proposed is $y = x/(1+x)$, since the transformation is monotone. Any other such transformation may be used but we propose to choose this transformation for its simplicity. Thus, the transformed sample which transforms the range to $(0,1)$ is given by $Y_i = X_i/(1+X_i)$ and the smoothed distribution is given by

$$\tilde{F}_n(x) = \tilde{G}_n(y), \quad \text{where } y = x/(1+x).$$

It is to be emphasized that even though any transformation of the random variable will be able to recover the original distribution, however for smooth estimator, this is not so. One simple reason is that the regions which have zero probability under the original distribution will carry some mass under the transformed scale. For example, if the transformation in item 2 is used when actually the distribution is concentrated on $[0,\infty)$, the smoothing weights will give some weight to $[-\infty,0)$ while the weights used in item 3 will give it zero weight.

For numerical illustration we obtain simulated samples from 4 distributions, (i) Beta$(2,4)$, (ii) Beta$(1,2)$, (iii) Exp(1) and (iv) $N(0,1)$. It is clear from the theorems that the choice of m is crucial in estimating the density as well as the distribution function. For the distribution function, m could be as large as $(n/\log n)^2$, but for the density estimation, it is desirable to have $m = o(n/\log n)$. Fig. 1 shows density estimates plotted for 10 simulated samples from Beta$(2,4)$ distribution for sample sizes $n = 30$, 60 and 125. For this figure, two values of m are contrasted, $m = n/\log n$ and $m = n$. One sees that for values of m as large as n, the variability in the density estimator increases. Various values of m were tried and we found that even though $m = o(n/\log n)$ is required, the choice $m = n/\log n$ is acceptable. This value of m is used in Figs. 2–5 for the four distributions, respectively. Each figure gives the plots for the new density estimator for sample sizes $n = 30$ and 125 for 10 simulated samples. These figures also show the plots of Kernel density estimators, using the `density` function of S-PLUS (1995), using the normal kernel and default window width. In general, the kernel method seems more variable and less accurate, especially when the value of the density is non-zero at a boundary (see Figs. 3 and 4). The solid line represents the true density in all the figures.

Fig. 6 gives the plot of smooth estimator of distribution functions for three distributions, Beta$(2,4)$, Exp(1) and $N(0,1)$. This figure gives the results on 10 simulated samples for the sample sizes $n = 30$ and 125. The new smooth estimator for the distribution function looks impressive in approximating the true distribution. Owing to the poor performance of the Kernel density estimator, we have not compared the corresponding estimator of the distribution function to the new estimator.

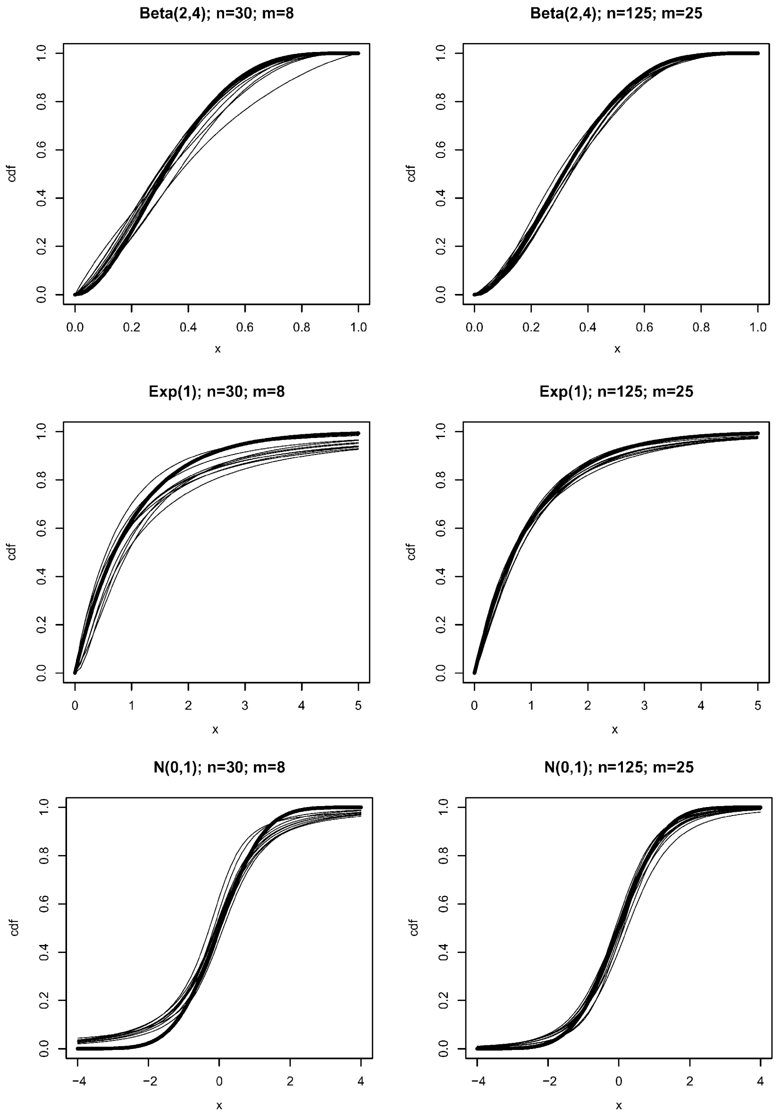

Fig. 6. Bernstein Polynomial Estimator for Distribution Function from Different Distributions.

Acknowledgements

This research work is partially supported by respective gresearch grants from NSERC of Canada, held by Angelo J. Canty and Yogendra P. Chaubey.

References

Babu, G.J., 1989. Strong representations for LAD estimators in linear models. Probab. Theory Related Fields 83, 547–558.

Babu, G.J., Singh, K., 1978. On deviations between empirical and quantile processes for mixing random variables. J. Multivariate Anal. 8, 532–549.

Chaubey, Y.P., Sen, P.K., 1996. On smooth estimation of survival and density functions. Statist. Decisions 14, 1–22.

Feller, W., 1965. An Introduction to Probability Theory and its Applications, Vol. II. Wiley, New York.

Härdle, W., 1991. Smoothing Techniques with Implementation in S. Springer, New York.

Parzen, E., 1962. On estimation of probability density and mode. Ann. Math Statist. 33, 1065–1070.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of density functions. Ann. Math Statist. 27, 832–837.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.

S-PLUS, 1995. S-PLUS Guide to Statistical and Mathematical Analysis. MathSoft Inc., Seattle.