Available online at www.sciencedirect.com
SCIENCE DIRECT®
Statistics & Probability Letters 76 (2006) 959–969
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors

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

[a]Department of Statistics, Pennsylvania State University, University Park, PA 16802-2111, USA
[b]Department of Mathematics and Statistics, Concordia University, Montreal, QC, Canada

Received 8 December 2004; received in revised form 14 September 2005
Available online 21 November 2005

Abstract

This paper considers multivariate extension of smooth estimator of the distribution and density function based on Bernstein polynomials studied in Babu et al. [2002. Application of Bernstein polynomials for smooth estimation of a distribution and density function. J. Statist. Plann. Inference 105, 377–392]. Multivariate version of Bernstein polynomials for approximating a bounded and continuous function is considered and adapted for smooth estimation of a distribution function concentrated on the hypercube $[0,1]^d$, $d>1$. The smoothness of the resulting estimator, naturally lends itself in a smooth estimator of the corresponding density. The functions with other compact or non-compact support can be dealt through suitable transformations. The asymptotic properties, namely, strong consistency and asymptotic normality of the resulting estimators are investigated under α-mixing. This has been motivated by estimation of conditional densities in non-linear dynamical systems.

MSC: primary 62G05; 62G07; 62G20

Keywords: Asymptotic normality; Kernel estimator; Non-parametric density estimation; Strong consistency; Strong mixing

1. Introduction

Let $\mathbf{X}$ denote a d-dimensional vector random variable with a distribution function denoted by F and a continuous density f, so that

$$F(\mathbf{x}) = \int_{\mathbf{t}\leqslant\mathbf{x}} f(\mathbf{t})\,\mathrm{d}\mathbf{t}. \tag{1.1}$$

(The ordered relations $\leqslant$ and $>$ are taken to be coordinate wise.) In this article, we shall implicitly consider the case $d>1$. Azzalini (1981) used the popular kernel estimator of a density for deriving the estimators of

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

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.10.031

quantiles and the distribution function. For bounded random variables this estimator, however, may assign positive mass to a zero probability region and hence, alternative estimators are desired to alleviate this situation. We refer to Scott (1992) for various methods of estimation of density that can also be used to derive an estimator of the distribution function, by integration. Chen (1999) introduced a beta kernel density estimator for random variables with compact support on $[0, 1]$, and recently Bouezmarni and Rolin (2003) have investigated its consistency properties.

Vitale (1973) introduced a Bernstein polynomial density function estimator which was later extended for estimation of densities concentrated on triangles and squares by Tenbusch (1994) for the i.i.d. case. On the other hand, motivated by the problem of smooth estimation of a continuous distribution function, Babu et al. (2002) obtained the same estimator as that of Vitale (1973) and investigated its large sample properties. In this paper we extend these results to multivariate case. We consider only distributions F with support $[0, 1]^d$; random variables with other compact or non-compact support can be transformed to the case considered here by using simple transformations as in Babu et al. (2002). These estimators are found to have comparable large sample properties to those given by the popular kernel method. Furthermore, they alleviate the usual deficiency of the kernel estimator that may assign non-zero probability to regions with zero probability.

The emphasis in this paper is on the density estimation for the case of multivariate mixing random variables, where as the estimator considered in Tenbusch (1994) is for the i.i.d. case. Tenbusch (1994) obtains the exact order of magnitude for the point wise mean squared error where as our results are on almost sure bounds for maximum difference between the estimator and the true value (i.e. L_∞ norm). The degree m of the Bernstein polynomials in our estimators and the sample size n are allowed to vary far more freely than the restrictions on Theorems 4a and b in Tenbusch (1994). Unlike Tenbusch (1994, Theorems 4a and b), we do not require the assumption $f(x, y) \neq 0$. In our case the assumption $f(x, y) > 0$ is used only to get non-zero variance in the CLT. Furthermore, we only assume Lipschitz continuity of order 1 where as Tenbusch (1994) assumes existence and boundedness of all partial derivatives of order 2.

The empirical distribution function F_n, based on a random sample $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ of n observations from the distribution F, is defined as

$$F_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} I_{(X_{ij} \leqslant x_j)} = \frac{1}{n}\sum_{i=1}^{n} I_{(X_{i1} \leqslant x_1, \ldots, X_{id} \leqslant x_d)}, \tag{1.2}$$

where X_{ij} denotes the jth component of the ith sample observation, and $\mathbf{x} = (x_1, x_2, \ldots, x_d)$.

The following is an extension of Theorem 1.1 of Babu et al. (2002) which can be easily established.

Theorem 1.1. *If u is a bounded and continuous function on the hypercube $[0, 1]^d$, then as $m \to \infty$*

$$u_m^*(\mathbf{x}) = \sum_{k_1=0}^{m}\sum_{k_2=0}^{m}\cdots\sum_{k_d=0}^{m} u\left(\frac{k_1}{m}, \frac{k_2}{m}, \ldots, \frac{k_d}{m}\right)\prod_{j=1}^{d} b_{k_j}(m, x_j) \to u(\mathbf{x}) \tag{1.3}$$

uniformly for $\mathbf{x} \in [0, 1]^d$, where

$$b_k(m, x) = \binom{m}{k} x^k (1-x)^{m-k}, \quad k = 0, \ldots, m. \tag{1.4}$$

In what follows, for brevity we consider the case $d = 2$, and write $X_i = X_{i1}$ and $Y_i = X_{i2}$ from now on. However the exposition easily extends to the case $d > 2$. For the bivariate case, Theorem 1.1 motivates the following smooth estimator for $F(x, y)$:

$$\tilde{F}_{n,m}(x, y) = \sum_{j=0}^{m}\sum_{k=0}^{m} F_n\left(\frac{j}{m}, \frac{k}{m}\right) b_j(m, x) b_k(m, y), \quad x, y \in [0, 1] \tag{1.5}$$

based on Bernstein polynomials. Note that $\tilde{F}_{n,m}$ is a polynomial in x and y, 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, $0 \leqslant \tilde{F}_{n,m}(x,y) \leqslant 1$ for $x \in [0,1]$, and

$$\tilde{F}_{n,m}(x,y) = \sum_{j=0}^{m} \sum_{k=0}^{m} f_n\left(\frac{j}{m},\frac{k}{m}\right) B_j(m,x) B_k(m,y), \tag{1.6}$$

where

$$f_n\left(\frac{j}{m},\frac{k}{m}\right) = F_n\left(\frac{j}{m},\frac{k}{m}\right) - F_n\left(\frac{j-1}{m},\frac{k}{m}\right) - F_n\left(\frac{j}{m},\frac{k-1}{m}\right) + F_n\left(\frac{j-1}{m},\frac{k-1}{m}\right) \tag{1.7}$$

for $j = 1,\ldots,m$; $k = 1,\ldots,m$, and for $i = 0,1,\ldots,m$, $f_n(0,i/m) = f_n(i/m,0) = 0$ and

$$B_i(m,x) = \sum_{j=i}^{m} b_j(m,x) = m\binom{m-1}{i-1} \int_0^x t^{i-1}(1-t)^{m-i}\,\mathrm{d}t. \tag{1.8}$$

Note that $f_n(j/m,k/m) \geqslant 0$ and $B_j(.,x)$ is non-decreasing in x, hence $\tilde{F}_{n,m}$ is a proper distribution function in R^2; that is, if

$$A = \{(x,y) \in R^2 : x_1 < x \leqslant x_2, y_1 < y \leqslant y_2\} \tag{1.9}$$

is a rectangular region in R^2, then

$$\begin{aligned}\tilde{F}_{n,m}(A) &= \tilde{F}_{n,m}(x_2,y_2) - \tilde{F}_{n,m}(x_1,y_2) - \tilde{F}_{n,m}(x_2,y_1) + \tilde{F}_{n,m}(x_1,y_1)\\ &= \sum_{j=0}^{m} \sum_{k=0}^{m} f_n\left(\frac{j}{m},\frac{k}{m}\right)(B_j(m,x_2) - B_j(m,x_1))(B_k(m,y_2) - B_k(m,y_1))\\ &\geqslant 0.\end{aligned}$$

Furthermore, the representation (1.6) leads to a smooth estimator

$$\begin{aligned}\tilde{f}_{n,m}(x,y) &= \sum_{j=1}^{m} \sum_{k=1}^{m} f_n\left(\frac{j}{m},\frac{k}{m}\right) \frac{\partial}{\partial x}\frac{\partial}{\partial y} B_j(m,x) B_k(m,y)\\ &= m^2 \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} f_n\left(\frac{(j+1)}{m},\frac{(k+1)}{m}\right) b_j(m-1,x) b_k(m-1,y)\end{aligned} \tag{1.10}$$

for density f of F. This problem has been motivated by the question of estimation of metric entropy of a non-linear dynamical system when the observations are contaminated by noise such as given by

$$x_{n+1} = \tau(x_n) + \xi_n,$$

where $x_n \in [0,1]$, $n = 0,1,\ldots$ and random variables ξ_n represent i.i.d. noise random variables. Under the assumption that τ admits an absolutely continuous invariant measure given by the density f_τ, it is useful to estimate f_τ and τ', the derivative of τ to obtain a plug-in estimator of metric entropy $h(\tau)$ (see Abarbanel, 1995) given by

$$h(\tau) = \int_0^1 \log|\tau'(x)| f_\tau(x)\,\mathrm{d}x.$$

Here, τ represents the regression of $x_{n+1}|x_n$, and f_τ represents the density of x_n, $n = 1,2,\ldots$. Such data may be considered to be subject to strong mixing (also called α-mixing). As such, we examine the asymptotic properties of $\tilde{F}_{n,m}$ and $\tilde{f}_{n,m}$, when the sequence $\{\mathbf{X}_1,\mathbf{X}_2,\ldots\}$ satisfies the following α-mixing condition.

Definition. A sequence $\{\mathbf{X}_i\}_{i=1}^{\infty}$ is called α-mixing (strong mixing) sequence if

$$|P(A \cap B) - P(A)P(B)| \leqslant \alpha(n)$$

for all $A \in \mathscr{M}^k$, $B \in \mathscr{M}_{n+k}$, $n \geqslant 1$ and all $k \geqslant 1$, where $\mathscr{M}^k$ denotes the σ-field generated by $(\ldots,\mathbf{X}_{k-1},\mathbf{X}_k)$ and $\mathscr{M}_k$ is the σ-field generated by $(\mathbf{X}_k,\mathbf{X}_{k+1},\ldots)$, and the sequence $\alpha(n)$ is a non-increasing sequence with $\lim_{n\to\infty}\alpha(n) = 0$.

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

Throughout this paper we use the notation

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

for a bounded function G on $\Omega = [0,1]^2$,

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

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 unit square Ω and $\sum_{i=1}^{\infty}\alpha(i)<\infty$. If $m,n\to\infty$, then $\|\tilde{F}_{n,m}-F\|\to 0$ a.s.*

Proof. We first note that for all $x,y\in[0,1]$,

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

Since $\|F_n-F\|\to 0$ a.s. as $n\to\infty$ (see Theorem 2.3 and Corollary 2.1 of Cai and Roussas, 1992) and F is uniformly continuous, the result follows from Theorem 1.1. □

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 $g\in R^2$ is said to satisfy Lipschitz condition of order $\gamma>0$, if there exists a constant C such that

$$|g(x,y)-g(x',y')|\leqslant C((x-x')^2+(y-y')^2)^{\gamma/2}. \tag{2.3}$$

Theorem 2.2. *Let $\alpha(i)=O(e^{-\theta i})$ for some $\theta>0$, and $n^{2/3}\leqslant m\leqslant n^{1-\varepsilon}$ for some fixed $0<\varepsilon\leqslant 1/3$. Let F be continuous and the first-order partial derivatives of F satisfy Lipschitz condition of order 1. Then as $n\to\infty$,*

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

To prove the theorem we need the following lemmas.

Lemma 2.1. *Let $\{Z_i\}$ be a strictly stationary α-mixing process with $|Z_i|\leqslant 1$, $E(Z_i)=0$ and $\alpha(n)=O(e^{-\theta n})$ for some $\theta>0$. Let $0<\delta<3/4$, $V>\mathrm{var}(Z_1)$ and $D<n^{3/4-\delta}V$. Then*

$$P\left(\left|\sum_{i=1}^{n} Z_i\right|>\rho D\right)\leqslant C(n^{-8}+\exp(-8(D^2/nV\log n))), \tag{2.5}$$

where $C>0$ and $\rho>0$ are constants depending only on δ and θ.

Proof. For any $r\geqslant 2$, as in Babu and Singh (1978), we have

$$\begin{aligned}\mathrm{var}\left(\sum_{i=1}^{r} Z_i\right) &\leqslant r\left(\mathrm{var}(Z_1)+\sum_{2\leqslant i\leqslant(2/\theta)\log r}|\mathrm{cov}(Z_1,Z_i)|+\sum_{i>(2/\theta)\log r}|\mathrm{cov}(Z_1,Z_i)|\right)\\ &\leqslant r\left(\mathrm{var}(Z_1)(1+(2/\theta)\log r)+4\sum_{i>(2/\theta)\log r} e^{-i\theta}\right)\\ &\leqslant rV(1+(2/\theta)\log r)+O(r^{-1}).\end{aligned} \tag{2.6}$$

Following the proof of Lemma 3.3 of Babu and Singh (1978), we obtain for any $y>0$, $\rho>0$, and an integer $A>0$, that

$$P\left(\left|\sum_{i=1}^{n} Z_i\right|>\rho D\right)\leqslant K_1\exp(-y\rho D+K_2y^2Vn\log n)+K_3n(y^4n)^A, \tag{2.7}$$

where K_1, K_2 depend only on θ, and K_3 depends only on θ and A. The result now follows from (2.6) and (2.7), by taking $\rho = \sqrt{32K_2}$, $y = \rho D(2K_2 Vn \log n)^{-1}$, and $A \geqslant 3/\delta$. $\square$

Lemma 2.2. *Let F satisfy Lipschitz condition of order 1, $\delta > 0$ and*

$$N_{x,m} = \{0 \leqslant k \leqslant m : |k - xm| \leqslant (m \log m)^{1/2}\}. \tag{2.8}$$

Then for $2 \leqslant m \leqslant n^{1-\delta}$, we have $H_{n,m} = \mathrm{O}(b_{n,m})$ a.s. as $n \to \infty$, where

$$H_{n,m} = \sup_{0<x,y<1} \max_{j \in N_{x,m}, k \in N_{y,m}} \left| F_n\left(\frac{j}{m}, \frac{k}{m}\right) - F\left(\frac{j}{m}, \frac{k}{m}\right) - F_n(x,y) + F(x,y) \right|. \tag{2.9}$$

Proof. Note that,

$$\begin{aligned} &\left| F_n\left(\frac{j}{m}, \frac{k}{m}\right) - F\left(\frac{j}{m}, \frac{k}{m}\right) - F_n(x,y) + F(x,y) \right| \\ &\quad \leqslant \left| F_n\left(\frac{j}{m}, \frac{k}{m}\right) - F\left(\frac{j}{m}, \frac{k}{m}\right) - F_n\left(\frac{j}{m}, y\right) + F\left(\frac{j}{m}, y\right) \right| + \left| F_n\left(\frac{j}{m}, y\right) - F_n(x,y) - F\left(\frac{j}{m}, y\right) + F(x,y) \right|, \end{aligned}$$

hence, we need to analyze the terms such as

$$|F_n(x',t) - F(x',t) - F_n(x,t) + F(x,t)|$$

and

$$|F_n(s,y') - F(s,y') - F_n(s,y) + F(s,y)|$$

for $|x'-x| \leqslant a_m$, $|y'-y| \leqslant a_m$, $s,t \in [0,1]$. We shall estimate the first term, second one can be estimated similarly. Clearly, for $|x'-s| \leqslant a_m$,

$$\begin{aligned} &|F_n(x',y) - F(x',y) - F_n(s,y) + F(s,y)| \\ &\quad \leqslant \max_{|i-j|b_{n,m} \leqslant 2a_m} |F_n(jb_{n,m}, y) - F_n(ib_{n,m}, y) - F(jb_{n,m}, y) + F(ib_{n,m}, y)| \\ &\qquad + 2\left[\max_{j \in N_{x,m}} |F((j+1)b_{n,m}, y) - F(jb_{n,m}, y)|\right] \\ &\quad \leqslant \max_{|i-j|b_{n,m} \leqslant 2a_m} |F_n(jb_{n,m}, y) - F_n(ib_{n,m}, y) - F(jb_{n,m}, y) + F(ib_{n,m}, y)| + \mathrm{O}(b_{n,m}), \end{aligned}$$

where the order of the second term in the above equation follows from the Lipschitz condition on F. Now, for $kb_{n,m} < y \leqslant (k+1)b_{n,m}$, using the monotonicity of F and F_n in each coordinate we get,

$$\begin{aligned} &F_n(jb_{n,m}, y) - F_n(ib_{n,m}, y) - F(jb_{n,m}, y) + F(ib_{n,m}, y) \\ &\quad \leqslant F_n(jb_{n,m}, (k+1)b_{n,m}) - F_n(ib_{n,m}, kb_{n,m}) - F(jb_{n,m}, kb_{n,m}) + F(ib_{n,m}, (k+1)b_{n,m}) \\ &\quad \leqslant F_n(jb_{n,m}, (k+1)b_{n,m}) - F_n(ib_{n,m}, kb_{n,m}) - F(jb_{n,m}, (k+1)b_{n,m}) + F(ib_{n,m}, kb_{n,m}) \\ &\qquad + F(jb_{n,m}, (k+1)b_{n,m}) - F(jb_{n,m}, kb_{n,m}) + F(ib_{n,m}, (k+1)b_{n,m}) - F(ib_{n,m}, kb_{n,m}) \\ &\quad \leqslant F_n(jb_{n,m}, (k+1)b_{n,m}) - F_n(ib_{n,m}, kb_{n,m}) - F(jb_{n,m}, (k+1)b_{n,m}) + F(ib_{n,m}, kb_{n,m}) + \mathrm{O}(b_{n,m}) \\ &\quad \leqslant F_n(jb_{n,m}, (k+1)b_{n,m}) - F_n(ib_{n,m}, (k+1)b_{n,m}) - F(jb_{n,m}, (k+1)b_{n,m}) + F(ib_{n,m}, (k+1)b_{n,m}) \\ &\qquad + F_n(ib_{n,m}, (k+1)b_{n,m}) - F_n(ib_{n,m}, kb_{n,m}) - F(ib_{n,m}, (k+1)b_{n,m}) + F(ib_{n,m}, kb_{n,m}) + \mathrm{O}(b_{n,m}). \end{aligned}$$

Together with a similar lower bound on the left-hand side expression, we get for $|i-j|b_{n,m} \leqslant 2a_m$,

$$|F_n(jb_{n,m}, y) - F_n(ib_{n,m}, y) - F(jb_{n,m}, y) + F(ib_{n,m}, y)| \leqslant D_{n,m,1} + D_{n,m,2} + \mathrm{O}(b_{n,m}), \tag{2.10}$$

where

$$D_{n,m,1} = \max_{|i-j|b_{n,m} \leqslant 2a_m} \max_{k \leqslant 1+b_{n,m}^{-1}} |F_n(jb_{n,m}, kb_{n,m}) - F_n(ib_{n,m}, kb_{n,m}) - F(jb_{n,m}, kb_{n,m}) + F(ib_{n,m}, kb_{n,m})|$$

and

$$D_{n,m,2} = \max_{|i-j|b_{n,m} \leqslant 2a_m} \max_{k \leqslant 1+b_{n,m}^{-1}} |F_n(kb_{n,m}, jb_{n,m}) - F_n(kb_{n,m}, ib_{n,m}) - F(kb_{n,m}, jb_{n,m}) + F(kb_{n,m}, ib_{n,m})|.$$

Therefore, it follows that

$$H_{n,m} \leqslant 2D_{n,m,1} + 2D_{n,m,2} + \mathrm{O}(b_{n,m}). \tag{2.11}$$

By applying Lemma 2.1 with $D = nb_{n,m}$ and $V = 2a_m$, we find a constant $\rho > 0$ such that for any y,

$$P(|F_n(jb_{n,m}, y) - F_n(ib_{n,m}, y) - (F(jb_{n,m}, y) - F(ib_{n,m}, y))| > \rho b_{n,m}) = \mathrm{O}(n^{-8}). \tag{2.12}$$

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

$$P(D_{n,m,1} > \rho b_{n,m}) = \mathrm{O}(n^2 n^{-8}) = \mathrm{O}(n^{-6}).$$

Therefore

$$\sum_n P(D_{n,m,1} > \rho b_{n,m}) < \infty.$$

By Borel–Cantelli lemma, we have $D_{n,m,1} = \mathrm{O}(b_{n,m})$ a.s. as $n \to \infty$. Similarly it follows that $D_{n,m,2} = \mathrm{O}(b_{n,m})$ a.s. as $n \to \infty$. Lemma 2.2 now follows from (2.11) $\square$

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

$$\tilde{F}_{n,m}(x, y) - F_n(x, y) = T_1(x, y) + T_2(x, y), \tag{2.13}$$

where

$$T_1(x, y) = (F_m^*(x, y) - F(x, y))$$

and

$$T_2(x, y) = \sum_{j=0}^{m} \sum_{k=0}^{m} b_j(m, x) b_k(m, y)(F_n(j/m, k/m) - F(j/m, k/m) - F_n(x, y) + F(x, y)).$$

Denote, the partial derivatives of F with respect to x, y as F_x and F_y, respectively. Since the first-order partial derivatives F_x and F_y of F satisfy Lipschitz condition of order 1, we have uniformly in $x, y \in (0, 1)$, that

$$\begin{aligned} F(j/m, k/m) - F(x, y) = {} & ((j/m) - x)F_x(x, y) + ((k/m) - y)F_y(x, y) \\ & + \mathrm{O}(((j/m) - x)^2 + ((k/m) - y)^2). \end{aligned} \tag{2.14}$$

Since,

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

it follows by (2.14) that

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

And therefore

$$\sup_{(x,y) \in \Omega} |T_1(x, y)| = \mathrm{O}(1/m). \tag{2.17}$$

Next, to investigate the second term of Eq. (2.13), we see that

$$\begin{aligned} |T_2(x, y)| & \leqslant \left(\left| \sum_{j \in N_{x,m}} \sum_{k \in N_{y,m}} b_j(m, x) b_k(m, y) a_{j,k,n,m}(x, y) \right| \right) + \left(2 \sum_{j \notin N_{x,m}} b_j(m, x) \right) + \left(2 \sum_{k \notin N_{y,m}} b_k(m, y) \right) \\ & = T_{21}(x, y) + T_{22}(x, y) + T_{23}(x, y), \end{aligned}$$

where

$$a_{j,k,n,m}(x,y) = F_n\left(\frac{j}{m},\frac{k}{m}\right) - F\left(\frac{j}{m},\frac{k}{m}\right) - F_n(x,y) + F(x,y). \tag{2.18}$$

It follows by applying Lemma 2.1 of Babu et al. (2002) (see Eq. (2.17)) 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 $s = 1/2$, we get for $m \geqslant 16$,

$$T_{22}(x,y) + T_{23}(x,y) \leqslant \frac{8}{m}. \tag{2.19}$$

Furthermore, since

$$T_{21}(x,y) \leqslant \sup_{0<s,t<1} \max_{j\in N_{s,m}} \max_{k\in N_{t,m}} |a_{j,k,n,m}(s,t)| = H_{n,m},$$

the theorem now follows from Lemma 2.2. □

Remark 2.1. Comparison of a smooth estimator of the distribution function and the empirical distribution function may be used in justifying the use of the former in Bootstrap technique. We may refer the reader to Shorack and Wellner (1986) (Chapter 23, Section 1) where a parallel result to Theorem 2.2 for kernel method for i.i.d. case appears (see Corollary 1, p. 764). This result depends on the specific kernel used, e.g. in the case of the Gaussian kernel, the optimum rate is $O(n^{-2/3})$, where as in our case the rate is close to $O(n^{-3/4})$. Further note that $\tilde{F}_{n,m}$ has all the derivatives, and is a highly smooth estimator of F. As the proof hinges on estimating the increments of $F_n - F$ using the ideas similar to Bahadur's (1966) representation of quantile, it is very likely that the rate given here is optimal, except possibly for logarithmic terms. The question of optimal rates is difficult to address and will be visited elsewhere.

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

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

$$f_0\left(\frac{j}{m},\frac{k}{m}\right) = F\left(\frac{j}{m},\frac{k}{m}\right) - F\left(\frac{j}{m},\frac{(k-1)}{m}\right) - F\left(\frac{(j-1)}{m},\frac{k}{m}\right) + F\left(\frac{(j-1)}{m},\frac{(k-1)}{m}\right), \tag{3.1}$$

$f_0(0,0) = 0$ and $f_0(0,k/m) = f_0(j/m,0) = 0$ for $j = 1,\dots,m;\ k = 1,\dots,m$. Note that

$$T_m(x,y) = \frac{\partial^2 F_m^*(x,y)}{\partial x \partial y} = m^2 \sum_{j=0}^{M}\sum_{k=0}^{M} f_0\left(\frac{j+1}{m},\frac{k+1}{m}\right) b_j(M,x) b_k(M,y) \tag{3.2}$$

is the density of the distribution F_m^*.

Theorem 3.1. *Assume that F has continuous density f, and that f satisfies Lipschitz condition of order* 1. *Let* $\alpha(i) = O(e^{-\theta i})$ *for some* $\theta > 0$. *For any* $\varepsilon > 0$ *we have as* $n \geqslant m \to \infty$,

$$\|T_m - f\| = O(m^{-1/2}) \quad \text{and} \quad \|\tilde{f}_{n,m} - T_m\| = O(mn^{-1/2}\log n + m^2 n^{-(3/4)+\varepsilon}) \quad a.s. \tag{3.3}$$

Consequently, if $m \leqslant n^{(3/8)-\varepsilon}$ *for some* $0 < \varepsilon < \frac{3}{8}$, *then* $\|\tilde{f}_{n,m} - f\| \to 0$ *a.s. as* $m, n \to \infty$.

Moreover, if the first partial derivatives of f satisfy Lipschitz condition of order $1/2$, *then* $\|T_m - f\| = O(m^{-3/4})$. *So in this case if we take* $m = n^{(1/4)-\varepsilon}$, *then*

$$\|\tilde{f}_{n,m} - f\| = O(n^{-(3/16)+(3/4)\varepsilon}).$$

Proof. We first estimate T_m. Since

$$m^2 f_0\left(\frac{j+1}{m},\frac{k+1}{m}\right) = m^2 \int_{j/m}^{(j+1)/m}\int_{k/m}^{(k+1)/m} f(s,t)\,dt\,ds \tag{3.4}$$

and f is assumed to satisfy Lipschitz condition of order 1, we have

$$m^2 f_0\left(\frac{j+1}{m},\frac{k+1}{m}\right) = f\left(\frac{j}{m},\frac{k}{m}\right) + \mathrm{O}(m^{-1}) \tag{3.5}$$

$$= f(x,y) + \mathrm{O}\left(\left|\frac{j}{m}-x\right| + \left|\frac{k}{m}-y\right|\right) + \mathrm{O}(m^{-1})$$

$$= f(x,y) + \mathrm{O}\left(\left|\frac{j}{M}-x\right| + \left|\frac{k}{M}-y\right|\right) + \mathrm{O}(m^{-1}) = \mathrm{O}(1) \tag{3.6}$$

uniformly in $j = 1,\ldots,m$; $k = 1,\ldots,m$, and in $0 \leqslant x, y \leqslant 1$. Hence using Cauchy–Schwartz inequality and (2.15), we have uniformly for $0 \leqslant x, y \leqslant 1$,

$$T_m(x,y) - f(x,y) = \mathrm{O}\left(\sum_{j=0}^{M}\left|\frac{j}{M}-x\right| b_j(M,x) + \sum_{k=0}^{M}\left|\frac{k}{M}-y\right| b_k(M,y)\right) + \mathrm{O}(m^{-1})$$

$$= \mathrm{O}(m^{-1/2}). \tag{3.7}$$

This establishes the first part of (3.3). To estimate $\|\tilde{f}_{n,m} - T_m\|$, define

$$W_i(x,y) = \sum_{j=0}^{M}\sum_{k=0}^{M}(I_{U(j,k;i,m)} - P(U(j,k;i,m)))b_j(M,x)b_k(M,y), \tag{3.8}$$

where

$$U(j,k;i,m) = \{j < mX_i \leqslant j+1, k < mY_i \leqslant k+1\}.$$

Clearly,

$$\tilde{f}_{n,m}(x,y) - T_m(x,y) = \frac{m^2}{n}\sum_{i=1}^{n} W_i(x,y), \tag{3.9}$$

$$P(U(j,k;i,m)) = f_0\left(\frac{j+1}{m},\frac{k+1}{m}\right), \tag{3.10}$$

and

$$nm^{-2}\|f_{n,m} - T_m\| \leqslant n \max_{0\leqslant j,k\leqslant m}\left|f_n\left(\frac{j+1}{m},\frac{k+1}{m}\right) - f_0\left(\frac{j+1}{m},\frac{k+1}{m}\right)\right|$$

$$\leqslant \max_{0\leqslant j,k\leqslant m}\left|\sum_{i=1}^{n}(I_{U(j,k;i,m)} - P(U(j,k;i,m)))\right|$$

Since for some constant $K > 1$

$$\mathrm{var}(I_{U(j,k;i,m)}) \leqslant P(U(j,k;i,m)) \leqslant Km^{-2}, \tag{3.11}$$

by taking $V = K\max(m^{-2}, n^{-(1/2)+2\varepsilon}(\log n)^{-2})$, $0 < \delta < \varepsilon$, and $D = (Vn)^{1/2}\log n$ in Lemma 2.1, we get for some constant $\rho > 0$,

$$P\left(\max_{0\leqslant j,k\leqslant m}\left|\sum_{i=1}^{n}(I_{U(j,k;i,m)} - P(U(j,k;i,m)))\right| \geqslant \rho(Vn)^{1/2}\log n\right) = \mathrm{O}(n^{-6}). \tag{3.12}$$

So the second part of (3.3) now follows by Borel–Cantelli lemma. The last part is an immediate consequence of Taylor expansion and the fact that mean and variance of (1.4) are mx and $mx(1-x)$. This completes the proof of Theorem 3.1. □

Remark 3.1. Under similar conditions (see Prakasa Rao (1983, Theorem 3.1.12)), the order of uniform convergence for the kernel estimator under the i.i.d. bivariate case is $\mathrm{O}(n^{-1/6}(\log\log n)^{1/2})$. This convergence rate holds in the weakly dependent case as well, due to a result of Neumann (1998). Neumann has proved that in the case of weakly dependent data, a sequence of independent random variables with the same marginals, along with the weakly dependent data can be constructed on some probability space such that

$$\|\hat{f}_h - \tilde{f}_h\| = \mathrm{O}(n^{-1/2}\log n),$$

where $\hat{f}_h$ denotes the kernel estimator of f based on the weakly dependent data and $\tilde{f}_h$ is the same based on the i.i.d. data. The convergence rate in our case is given by $\mathrm{O}(n^{-(3/16)+\varepsilon)})$, much smaller than $\mathrm{O}(n^{-1/6})$, when ε is chosen to be small. This compares favorably to the convergence rate in the case of kernel estimator. The rate is improved to $\mathrm{O}(n^{-(1/4)+\varepsilon)})$, if Lipschitz condition of order 1 is assumed for the first-order partial derivatives of f.

For the rest of the paper we assume that, $\alpha(i) = \mathrm{O}(\mathrm{e}^{-\theta i})$, for some $\theta > 0$ and fix $0 < x, y < 1$. To establish asymptotic normality of $\tilde{f}_{n,m}(x,y)$, let

$$\sigma^2(x,y) = \lim_{m\to\infty} m^3(\mathrm{Var}(W_1(x,y)) + 2\sum_{i=2}^{\infty}\mathrm{Cov}(W_1(x,y), W_i(x,y))) > 0. \tag{3.13}$$

It is a simple algebra to establish that

$$\sigma^2_{r,m} = \sigma^2_{r,m}(x,y) = m^3 r^{-1}\mathrm{Var}\left(\sum_{i=1}^{r} W_i(x,y)\right) \to \sigma^2(x,y), \tag{3.14}$$

as $r > (\log m)^3 \to \infty$. Clearly, $\sum_{i=2}^{\infty}|\mathrm{Cov}(W_1(x,y), W_i(x,y))| < \infty$, whenever $\sum_{i=1}^{\infty}\alpha(i) < \infty$. By Lemma 3.1 of Babu et al. (2002), we have $m^3\mathrm{Var}(W_1) \to \gamma(x,y)$, where

$$\gamma(x,y) = (f(x,y)/4\pi)(xy(1-x)(1-y))^{-1/2} > 0, \tag{3.15}$$

provided $f(x,y) > 0$.

Remark 3.2. If $\{(X_i, Y_i)\}_{i=1}^{\infty}$ is a sequence of independent random vectors, then $\sigma^2(x,y) = \gamma(x,y)$. If the joint densities g_i of (X_1, Y_1, X_i, Y_i) are uniformly bounded then also we have $\sigma^2(x,y) = \gamma(x,y)$. This holds, in particular, when the sequence $\{(X_i, Y_i)\}$ is s-dependent and each of g_i $(2 \leqslant i \leqslant s)$ are bounded.

The next theorem is of interest on its own. However a consequence of it, stated as Corollary 1, gives point wise convergence rate of an estimator of $f(x,y)$. To state the theorem on asymptotic normality, note that from (3.7) and (3.8), for $0 < x, y < 1$,

$$|W_i(x,y)| \leqslant 1 \quad \text{and} \quad \sup_i |W_i(x,y)| \leqslant \max_{0\leqslant j,k\leqslant M} b_j(M,x)b_k(M,y) + \|T_m\| m^{-2} = \mathrm{O}(m^{-1}). \tag{3.16}$$

Theorem 3.2. *Assume that F has continuous density f, and that f satisfies Lipschitz condition of order* 1. *Let $\alpha(i) = \mathrm{O}(\mathrm{e}^{-\theta i})$ for some $\theta > 0$. If $3 \leqslant m \leqslant n^{(1/4)-\varepsilon}$ for some $0 < \varepsilon < \frac{1}{4}$, then*

$$(n/m)^{1/2}(\tilde{f}_{n,m}(x,y) - T_m(x,y)) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \sigma^2(x,y)) \quad \text{as } n \to \infty. \tag{3.17}$$

Proof. Let $[s]$ denote the largest integer not exceeding s. Following the proof of Theorem 27.4 of Billingsley (1995, pp. 364–367), define $\beta = \frac{1}{2} + 2\varepsilon$, $\alpha = \frac{1}{2} + \varepsilon$, $p = [n^\beta]$, $q = [n^\alpha]$, $r = [(n-p)/(p+q)] + 1$,

$$Z_i = m^{3/2} W_i(x,y), \quad S_\ell = \sum_{i=1}^{\ell} Z_i, \quad S_0 = 0$$

and write

$$(n/m)^{1/2}(\tilde{f}_{n,m}(x,y) - T_m(x,y)) = S_n = \sum_{i=1}^{r} U_i + \sum_{i=1}^{r} V_i,$$

where for $1 \leqslant i \leqslant r$,

$$V_i = S_{\min\{n, i(p+q)\}} - S_{(i-1)(p+q)+p},$$

$$U_i = S_{(i-1)(p+q)+p} - S_{(i-1)(p+q)}.$$

As in the proof of Theorem 27.4 of Billingsley (1995, p. 367), $(rp)^{-1/2}\sum_{i=1}^{r} U_i$ and $(rp)^{-1/2}\sum_{i=1}^{r} U_i'$ have the same limiting distribution, where U_i' are independent random variables with the common distribution same as that of U_i. As $(rp/n) \to 1$, it follows by Lindeberg Central Limit Theorem that

$(rp)^{-1/2}\sum_{i=1}^{r} U'_i \xrightarrow{\mathscr{D}} \mathcal{N}(0,\sigma^2(x,y))$, provided for each $\xi>0$,

$$\frac{1}{p}E(U_1^2 I_{(|U_1|>\xi\sqrt{n})}) \to 0. \tag{3.18}$$

By (3.16), we have

$$U_1^2 \leqslant p^2 m^3 \sup_i |W_i(x,y)|^2 = \mathrm{O}(p^2 m),$$

$$\begin{aligned}|U_1| &= m^{3/2}\left|\sum_{i=1}^{p} W_i(x,y)\right| \\ &\leqslant m^{3/2} \max_{0\leqslant j,k\leqslant m}\left|\sum_{i=1}^{p}(I_{U(j,k;i,m)} - P(U(j,k;i,m)))\right|.\end{aligned}$$

Hence by (3.12),

$$\begin{aligned}\frac{1}{p}E(U_1^2 I_{(|U_1|>\xi\sqrt{n})}) &= \mathrm{O}\left(pmP\left(\max_{0\leqslant j,k\leqslant m}\left|\sum_{i=1}^{p}(I_{U(j,k;i,m)} - P(U(j,k;i,m)))\right| \geqslant \rho(Vp)^{1/2}\log p\right)\right) \\ &= \mathrm{O}(mp^{-5}) \to 0,\end{aligned}$$

provided

$$m^{3/2}(Vp/n)^{1/2}\log p \to 0, \tag{3.19}$$

where

$$V = K\max(m^{-2},\, p^{-(1/2)+2\varepsilon}(\log p)^{-2})$$

for some $K>1$. Clearly (3.19) holds if $m<n^{(1/4)-\varepsilon}$. This proves (3.18) and hence

$$n^{-1/2}\sum_{i=1}^{r} U_i \xrightarrow{\mathscr{D}} \mathcal{N}(0,\sigma^2(x,y)).$$

Similarly,

$$(rq)^{-1/2}\sum_{i=1}^{r} V_i \xrightarrow{\mathscr{D}} \mathcal{N}(0,\sigma^2(x,y)).$$

As $(rq/n)\to 0$, we conclude that $n^{-1/2}\sum_{i=1}^{r} V_i \to 0$ in probability and consequently,

$$n^{-1/2}S_n \xrightarrow{\mathscr{D}} \mathcal{N}(0,\sigma^2(x,y)).$$

This completes the proof of Theorem 3.2. □

Remark 3.3. If $mn^{-1/5}\to 0$ and

$$T_m(x,y) = f(x,y) + \frac{1}{m}\eta(x,y) + \mathrm{O}(m^{-2}). \tag{3.20}$$

For some function $\eta(x,y)$, then under the conditions of Theorem 3.2 we have

$$(n/m)^{1/2}(\tilde{f}_{n,m}(x,y) - f(x,y) + m^{-1}\eta(x,y)) \xrightarrow{\mathscr{D}} \mathcal{N}(0,\sigma^2(x,y)). \tag{3.21}$$

Therefore we have the following corollary.

Corollary 1. *In addition to the conditions of Theorem* 3.2, *suppose* $mn^{-1/5}\to 0$, *and the third-order partial derivatives of f satisfy Lipschitz condition of order* 1, *then* (3.21) *holds, with*

$$\eta(x,y) = (1-2x)f_x(x,y) + (1-2y)f_y(x,y) + x(1-x)f_{xx}(x,y) + y(1-y)f_{yy}(x,y), \tag{3.22}$$

where f_x, f_y *and* f_{xx}, f_{xy}, f_{yy} *denote the first- and second-order partial derivatives of f.*

Proof. By (3.4) and Taylor series expansion, we have,

$$
\begin{aligned}
m^2 f_0\left(\frac{j+1}{m}, \frac{k+1}{m}\right) &= f\left(\frac{j}{m}, \frac{k}{m}\right) + m \int_{j/m}^{(j+1)/m} \left(s - \frac{j}{m}\right) f_x\left(\frac{j}{m}, \frac{k}{m}\right) \mathrm{d}s \\
&\quad + m \int_{k/m}^{(k+1)/m} \left(t - \frac{k}{m}\right) f_y\left(\frac{j}{m}, \frac{k}{m}\right) \mathrm{d}t + \mathrm{O}(m^{-2}) \\
&= f\left(\frac{j}{m}, \frac{k}{m}\right) + \frac{1}{2m}\left(f_x\left(\frac{j}{m}, \frac{k}{m}\right) + f_y\left(\frac{j}{m}, \frac{k}{m}\right)\right) + \mathrm{O}(m^{-2}).
\end{aligned} \tag{3.23}
$$

A similar expansion yields that

$$
f_x\left(\frac{j}{m}, \frac{k}{m}\right) = f_x(x, y) + \left(\frac{j}{m} - x\right) f_{xx} + \left(\frac{k}{m} - y\right) f_{xy} + \mathrm{O}(m^{-1}).
$$

This leads to

$$
\sum_{j=0}^{M} \sum_{k=0}^{M} f_x\left(\frac{j}{m}, \frac{k}{m}\right) b_j(M; x) b_k(M; y) = f_x(x, y) + \mathrm{O}(m^{-1}).
$$

Repeated applications of such Taylor series expansions establish (3.20) with η given by (3.22). This proves the corollary. □

Remark 3.4. The above corollary gives the rate $\mathrm{O}(n^{-2/5})$ of convergence to normality after bias correction which compares favorably to the rate $\mathrm{O}(n^{-1/3})$ of kernel estimator in the bivariate i.i.d. case. This can be obtained by using Theorem 3.1.15 of Prakasa Rao (1983) with $h_n = n^{-1/6}$.

Acknowledgements

This research work is supported in part by Yogendra P. Chaubey's NSERC of Canada Grant, and G. Jogesh Babu's National Science Foundation Grant AST-0434234. The authors would also like to thank a referee whose suggestions have resulted in an improved version of the paper.

References

Abarbanel, H., 1995. Analysis of Observed Chaotic Data. Springer, New York.
Azzalini, A., 1981. A note on the estimation of a distribution function and quantiles. Biometrika 68, 326–328.
Babu, G.J., Singh, K., 1978. On deviations between empirical and quantile processes for mixing random variables. J. Multivariate Anal. 8, 532–549.
Babu, G.J., Canty, A.J., Chaubey, Y.P., 2002. Application of Bernstein polynomials for smooth estimation of a distribution and density function. J. Statist. Plann. Inference 105, 377–392.
Bahadur, R.R., 1966. A note on quantiles in large samples. Ann. Math. Statist. 37, 577–580.
Billingsley, P., 1995. Probability and Measure, third ed. Wiley, New York.
Bouezmarni, T., Rolin, J.-M., 2003. Consistency of the beta kernel density function estimator. Canad. J. Statist. 31, 89–98.
Cai, Z., Roussas, G.G., 1992. Uniform strong estimation under α-mixing, with rates. Statist. Probab. Lett. 15, 47–55.
Chen, S.X., 1999. Beta kernel estimators for density functions. Comput. Statist. Data Anal. 31, 131–145.
Neumann, M.H., 1998. Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. Ann. Statist. 26, 2014–2048.
Prakasa Rao, B.L.S., 1983. Nonparametric Functional Estimation. Academic Press, London.
Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice and Visualization. Wiley, New York.
Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.
Tenbusch, A., 1994. Two-dimensional Bernstein polynomial density estimators. Metrika 41, 233–253.
Vitale, R.A., 1973. A Bernstein polynomial approach to density estimation. Comm. Statist. 2, 493–506.