Smooth estimation of multivariate survival and density functions

Yogendra P. Chaubey*, Pranab K. Sen
Concordia University, Montreal, Quebec, Canada and University of North Carolina, Chapel Hill, NC, USA

Received 17 December 1999; accepted 19 December 1999

Abstract

By considering a multivariate version of the Hille’s theorem, the technique developed by Chaubey and Sen (Statist. Decisions 14 (1996) 1) is extended for estimating a multivariate survival distribution and its associated density. Asymptotic properties of the resulting estimators are investigated in detail. © 2002 Elsevier Science B.V. All rights reserved.

MSC: 62G05; 62G07; 62G20

Keywords: Bahadur representation; Exponential smoothing; Hille’s theorem; Strong consistency; Survival function

1. Introduction

Let $X$ denote a $d$-dimensional vector random variable defined on $\mathbb{R}^d$, with a distribution function denoted by $F(x)$ and a continuous density $f(\cdot)$, so that

$$F(x) = \int_{t \leq x} f(t) \, dt.$$  \hfill (1.1)

The corresponding multivariate survival function is defined as

$$S(x) = \int_{t > x} f(t) \, dt.$$  \hfill (1.2)

(The ordered relations $\leq$ and $>$ are taken to be coordinate wise.) These distributions are quite common in biomedical and industrial studies where each component of the random vector $X$ represents multivariate failure time of a system which survives as long as any one unit of the system survives (see Cox and Oakes, 1984). Note that in the univariate case $F(x) + S(x) = 1$ for all $x$, however, this may not be true in the...
multivariate case. Thus, estimation of $S(x)$ does not follow directly from that of $F(x)$ in general for $d > 1$ as opposed to the case $d = 1$. In this article, we shall implicitly consider the case $d > 1$. Based on a random sample $(X_1, X_2, \ldots, X_n)$ of $n$ failure times of the system, $F(x)$ is, generally, estimated by the empirical distribution function, $F_n(x)$, defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} I(X_{ij} \leq x_j) = \frac{1}{n} \sum_{i=1}^{n} I(X_{i1} \leq x_1, \ldots, X_{id} \leq x_d), \quad (1.3)$$

where $X_{ij}$ denotes the $j$th component of the $i$th sample observation. The corresponding estimator of the survival function $S(x)$ is given by

$$S_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_{i1} > x_1, \ldots, X_{id} > x_d). \quad (1.4)$$

The discontinuities in $F_n(x)$ and $S_n(x)$ render these estimators useless as “graphical tools” and otherwise in providing smooth estimators of the density function. As a result a lot of effort has been afforded to smooth estimation of multivariate density functions (see Scott, 1992).

For $d = 1$, the pioneering work of Rosenblatt (1956) and Parzen (1962) created a tremendous scope for smooth estimation of density and related functionals. The offshoot of this approach in the area of nonparametric regression has rendered this area an enormous vitality due to its wide ranging applications in health sciences and communications. We may refer to some recent excellent monographs highlighting the details of these methods, e.g. Fan and Gijbels (1996), Wand and Jones (1995), Green and Silverman (1994) and Härdle (1991). Some of the univariate methods have been extended to the multivariate case (see for example, Cacoullos, 1966; Deheuvels, 1977; Loftsgaarden and Quesenberry, 1965; Murthy, 1966). Some of these methods can be treated under a unified manner using the so called “delta-sequence” (see Bosq, 1977; Földes and Révész, 1974; Walter and Blum, 1979). Stone’s (1977) paper can be also put in this category with respect to the nonparametric regression. Susarla and Walter (1981) extended it to the multivariate case.

The popularity of the kernel estimator for density estimation in the univariate case also extends to the multivariate case. One reason for this is the generality of the kernel method as proved by Terrell and Scott (1992) which showed that many smooth estimators can be expressed as a kernel estimator. But, other methods such as multivariate splines (see Chui, 1988) and $k$-NN ($k$ nearest neighbor) (see Hall, 1983, 1985) methods have also been extensively investigated.

The methods mentioned above are general in nature and do not account for specific feature of the data defined on the positive orthant and hence may assign positive probability to a region which should have zero probability. Our objective here is to extend the methodology developed in Chaubey and Sen (1996) for smooth estimation of the distribution and density functions in the univariate case. Section 2 gives the basic motivation along with the proposed estimator of the survival function, distribution function and the density function and Section 3 discusses their properties.
2. Preliminary notions

Here, we generalize the estimator of Chaubey and Sen (1996) to the multivariate case (however, here we consider the untruncated case only). The smooth estimator of the survival function $S(x)$ is given by

$$
\hat{S}_n(x) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \cdots \sum_{j_d=0}^{\infty} w_{j_1,\ldots,j_d}(x_1,\ldots,x_d) S_n \left( \frac{j_1}{\lambda_{1n}}, \ldots, \frac{j_d}{\lambda_{dn}} \right),
$$

(2.1)

where the weights $w_{j_1,\ldots,j_d}(t_1,\ldots,t_d)$ are defined as

$$
w_{j_1,\ldots,j_d}(t_1,\ldots,t_d) = e^{-\sum_{i=1}^{d} t_i} \prod_{i=1}^{d} \left( \frac{t_i}{j_i} \right)!
$$

(2.2)

and $\lambda_{in}$ may be chosen as

$$
\lambda_{in} = \frac{n}{\max(X_{i1},\ldots,X_{in})}, \quad i = 1, 2, \ldots, d.
$$

(2.3)

The above choice of the constant $\lambda_{in}$ is obviously stochastic and is motivated from the fact that there is no observed mass beyond the point $\max(X_{i1},\ldots,X_{in})$ for the variable $X_i$. The deterministic choice has to be such that $\lambda_{in} \to \infty$ but $n^{-1} \lambda_{in} \to 0$ (see Chaubey and Sen, 1996 for discussion on other choices).

In the univariate case the positive half of the real line is partitioned into a lattice of points $0, 1, 2, \ldots$ and a suitable Poisson distribution is superimposed on the empirical survival function to provide the smooth estimator. In the multivariate case, we consider a lattice of hypercubes and generate the weights by product of proper Poisson weights. The corresponding estimator of the distribution function is given as

$$
\hat{F}_n(x) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \cdots \sum_{j_d=0}^{\infty} w_{j_1,\ldots,j_d}(x_1,\ldots,x_d) F_n \left( \frac{j_1}{\lambda_{1n}}, \ldots, \frac{j_d}{\lambda_{dn}} \right).
$$

(2.4)

The smooth estimator of the density function may be readily obtained by considering the derivative of $\hat{F}_n(x)$ which is given by

$$
\hat{f}_n(x) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \cdots \sum_{j_d=0}^{\infty} \frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} w_{j_1,\ldots,j_d}(x_1,\ldots,x_d) F_n \left( \frac{j_1}{\lambda_{1n}}, \ldots, \frac{j_d}{\lambda_{dn}} \right).
$$

(2.5)

The above expression for the density estimator is appropriate for computational purpose, however, it may not be useful for studying its theoretical properties. As we will
concentrate on the bivariate distributions for further exposition, an alternative form will be provided which would be useful for theoretical investigation (see below). The following simpler notations will be used for the bivariate case where the bivariate random variable is denoted by \((X, Y)\) and the corresponding estimators in obvious notations are given as

\[
\hat{S}_n(x, y) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}(\lambda_1 n x, \lambda_2 n y) S_n \left( \frac{j}{\lambda_1 n}, \frac{k}{\lambda_2 n} \right),
\]

(2.6)

\[
\hat{F}_n(x, y) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}(\lambda_1 n x, \lambda_2 n y) F_n \left( \frac{j}{\lambda_1 n}, \frac{k}{\lambda_2 n} \right),
\]

(2.7)

and

\[
\hat{f}_n(x, y) = \lambda_1 n \lambda_2 n \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}(\lambda_1 n x, \lambda_2 n y) \left[ F_n \left( \frac{j}{\lambda_1 n}, \frac{k}{\lambda_2 n} \right) - F_n \left( \frac{j+1}{\lambda_1 n}, \frac{k}{\lambda_2 n} \right) + F_n \left( \frac{j+1}{\lambda_1 n}, \frac{k+1}{\lambda_2 n} \right) - F_n \left( \frac{j}{\lambda_1 n}, \frac{k+1}{\lambda_2 n} \right) \right],
\]

(2.8)

where

\[
w_{jk}(t_1, t_2) = e^{-(t_1 + t_2)} \frac{t_1^j t_2^k}{j! k!}
\]

(2.9)

with \(\lambda_1 n\) and \(\lambda_2 n\) being given by

\[
\lambda_1 n = \frac{n}{\max(X_1, X_2, \ldots, X_n)}, \quad \lambda_2 n = \frac{n}{\max(Y_1, Y_2, \ldots, Y_n)}.
\]

(2.10)

The motivation of the estimators presented here comes from the following multivariate extension of the Hille’s theorem (see Feller, 1965, p. 219).

**Lemma 2.1.** Consider a sequence \(\{\Phi_n(y, t)\}_{n=1}^{\infty}\) of distribution functions in \(\mathbb{R}^d\) for every fixed \(t \in \mathbb{R}^d\), such that for \(Y_n \sim \Phi_n(\cdot, t)\)

(i) \(EY_n = t\)

(ii) \(v_n(t) = \max_{1 \leq i \leq d} \text{Var}(Y_{i,n}) \rightarrow 0\) as \(n \rightarrow \infty\), for every fixed \(t\).

(iii) \(\Phi_n(y, t)\) is continuous in \(t\).

Define for any bounded continuous multivariate function \(u(t)\)

\[
\tilde{u}_n(t) = \int_{\mathbb{R}^d} u(x) d\Phi_n(x, t).
\]

(2.11)

Then \(\tilde{u}_n(t) \rightarrow u(t)\) as \(n \rightarrow \infty\) for \(t\) in any compact subset of \(\mathbb{R}^d\), the convergence being uniform over any subset over which \(u(t)\) is uniformly continuous. Furthermore, if the function \(u(t)\) is monotone, the convergence holds uniformly over entire \(\mathbb{R}^d\).
Proof. The proof directly follows from the celebrated Helly–Bray theorem (see Theorem 3.2.2 in Sen and Singer, 1993) as the distribution $\Phi_n(y, t)$ converges to a degenerate distribution at the point $t$ under the stated conditions.

Choosing $\Phi_n$ such that it gives mass $\prod_{i=1}^d e^{-\frac{t_i}{\lambda_i}}$ to the point $(\frac{j_1}{\lambda_1}, \ldots, \frac{j_d}{\lambda_d})$, proves that

$$
\sup_{t \in \mathbb{R}^d} |S_n^*(t) - S(t)| \to 0 \quad \text{as } n \to \infty, \quad (2.12)
$$

where

$$
S_n^*(x) = \sum_{j_1=0}^\infty \sum_{j_2=0}^\infty \cdots \sum_{j_d=1}^\infty w_{j_1, \ldots, j_d} S\left(\frac{j_1}{\lambda_1}, \ldots, \frac{j_d}{\lambda_d}\right). \quad (2.13)
$$

The stochastic version of the above result is obtained upon replacing $S(j_1/\lambda_1, \ldots, j_d/\lambda_d)$ by its empirical version $S_n(j_1/\lambda_1, \ldots, j_d/\lambda_d)$. The main results on the asymptotic properties of these estimators are given in the following section.

Remark 2.1. A natural interest is also to explore the practical case of censored survival data. Since the components of the random vector $X$ are generally not independent, the mechanism of censoring may be much more complex than in the univariate case. A censoring may be with respect to one response variable only; for example, $X_1$ relates to the primary end point, while the remaining $d - 1$ coordinates refer to secondary or surrogate end points, a censoring may specifically relate to $X_1$ alone. In such a case, under noninformative and random censoring scheme, we may conceive of a nonnegative random variable $C$ which is independent of $X_1$ (and does not depend on the other auxiliary or concomitant variables), such that the observable random elements are (i) $X_1^o = \min(X_1, C)$, (ii) $\delta = I(X_1^o = X_1)$, and (iii) $X_j, j = 2, \ldots, d$, along with other auxiliary variables. In the more likely case, a censoring may be due to withdrawal or drop-out, and if so, it automatically removes the observation vector as a whole from further study. In that case, we have observed multivariate survival time $X^o = (\min(X_j, C), j = 1, \ldots, d)$. The well known Kaplan–Meier (KM) product limit estimator of the survival function in the univariate case has been extended to specific multivariate models under specific regularity assumptions (see Dabrowska, 1988, 1989a, b; Dabrowska et al., 1999; Pedroso de Lima and Sen, 1997, 1999 for some discussion). Given these developments, one can incorporate the univariate results of Chaubey and Sen (1998) on smoothing of the Product Limit estimator, and formulate their multivariate counterparts. For intended brevity of presentation, we omit the details. It is also worthwhile to note that Dabrowska’s results relate to the KM estimator under additional conditions that are not likely to be tenable in all practical applications. Likewise, the counting process approach inherits some multiplicative intensity functions which may not also be universally true. From these perspectives, it seems that there is ample room for the appraisal of censoring schemes in the multivariate case, and in the light of this
assessment an appropriate scheme should be chosen. Whatever be the outcome, the smoothing procedure considered here remains adoptable.

3. Main results

3.1. Properties of the smooth estimator of the survival and distribution function

We may now extend the theorem in Chaubey and Sen (1996) on the strong consistency of the estimator $\tilde{F}_n(t)$.

**Theorem 3.1.** Suppose that for each $i$, $\lambda_{in} \to \infty$, $n^{-1}\lambda_{in} \to 0$ as $n \to \infty$ and $f(x)$ has bounded first derivatives then

$$
\|\tilde{F}_n - F\| = \sup_{x \in \mathbb{R}^+} |\tilde{F}_n(x) - F(x)| \to 0 \text{ a.s. as } n \to \infty.
$$

(3.1)

**Proof.** Let

$$
F^*_n(x) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \cdots \sum_{j_d=0}^{\infty} w_{j_1,\ldots,j_d}(x_1^{\lambda_{1n}},\ldots,x_d^{\lambda_{dn}}) F\left( \frac{j_1}{\lambda_{1n}}, \ldots, \frac{j_d}{\lambda_{dn}} \right).
$$

(3.2)

Since $F(x)$ is a bounded and absolutely continuous, for any compact subset $C$ of $\mathbb{R}^+$, by the use of Lemma 2.1, we have

$$
\sup_{x \in C} |F^*_n(x) - F(x)| \to 0 \text{ a.s. as } n \to \infty.
$$

(3.3)

for $\lambda_{in} \to \infty$, $i = 1,2,\ldots,d$. Now since

$$
\sup_{x \in C} |\tilde{F}_n(x) - F(x)| \leq \sup_{x \in C} |\tilde{F}_n(x) - F^*_n(x)| + \sup_{x \in C} |F^*_n(x) - F(x)|,
$$

(3.4)

it is easy to see that

$$
\sup_{x \in C} |\tilde{F}_n(x) - F^*_n(x)| \leq \max_{j_1,j_2,\ldots,j_d} \left| F_n\left( \frac{j_1}{\lambda_{1n}}, \ldots, \frac{j_d}{\lambda_{dn}} \right) - F\left( \frac{j_1}{\lambda_{1n}}, \ldots, \frac{j_d}{\lambda_{dn}} \right) \right|
$$

$$
\leq \sup_{x \in \mathbb{R}^+} |F_n(x) - F(x)|.
$$

(3.5)

By the multivariate version of the Glivenko–Cantelli lemma (see Sen and Singer, 1993), the right-hand side of (3.5) converges to zero a.s. as $n \to \infty$. Thus, using Eqs. (3.3)–(3.5) we can claim that for any compact set $C$, $\sup_{x \in C} |\tilde{F}_n(x) - F(x)|$ converges to zero a.s. as $n \to \infty$. For any component of $x$ going to $\infty$, since $\tilde{F}_n(x)$ and $F(x)$ are bounded and monotone in each coordinate, by Eq. (3.4) the difference can be arbitrarily bounded for sufficiently large $n$. This completes the proof of theorem. □
Theorem 3.2. Suppose that $F$ is absolutely continuous and for each $i$, $\lambda_i \to \infty$ and $n^{-1}\lambda_i \to 0$ then we have for any $\eta > 0$

$$\sup_{x \in \mathbb{R}^d} |\tilde{F}_n(x) - F_n(x)| = O(n^{-3/4}(\log n)^{1+\eta}) \quad a.s. \quad (3.6)$$

Before proving the theorem, first we generalize Lemma 1 of Bahadur (1966) to the multivariate case.

**Lemma 3.1.** Consider a sequence $a_n = n^{-1/2}(\log n)^{1+\eta}$, $\eta > 0$, $n = 1, \ldots, \infty$ and let $D_n(\xi) = \{x : |x_i - \xi_i| < a_n, \ i = 1, \ldots, d\}$ then for $G_n(x, \xi) = [F_n(x) - F_n(\xi)] - [F(x) - F(\xi)]$ we have

$$\sup_{\xi} \sup_{x \in D_n(\xi)} |G_n(x)| = O(n^{-3/4}(\log n)^{1+\eta}). \quad (3.7)$$

**Proof.** We provide the proof for the bivariate case only, the general case follows similarly. We need to consider the case for $\xi \in [0, 1] \times [0, 1]$ only, because with $W = F(X, \infty)$ and $Z = F(\infty, Y)$, we have $F_n(x, y) - F(x, y) = Q_n(w, z) - Q(w, z)$, where $Q$ and $Q_n$ are counterparts of $F$ and $F_n$, respectively, for $(W, Z)$, with $w = F(x, \infty)$ and $z = F(\infty, y)$, $(w, z) \in [0, 1] \times [0, 1]$. Let us write the set $D_n = D_n(\xi)$ as $(L_1, U_1) \times (L_2, U_2)$, then for a fixed $\xi = (\xi_1, \xi_2)$ we have $L_1 = \xi_1 - a_n, L_2 = \xi_2 - a_n, U_1 = \xi_1 + a_n, U_2 = \xi_2 + a_n$. Partition the set $D_n$ into $D_{njk} = (L_{nj}^{(1)}, L_{nj}^{(1)}) \times (L_{nk}^{(2)}, L_{nk}^{(2)})$ for $|j| \leq b_n, |k| \leq b_n$ where $L_{nj}^{(1)}, L_{nk}^{(2)}$, $i = 1, 2$ denote the subdivisions of $(L_1, U_1)$ and $(L_2, U_2)$ respectively, i.e. $L_{nj}^{(1)} = \xi_1 + ja_n b_n^{-1}$ and $L_{nk}^{(2)} = \xi_2 + ka_n b_n^{-1}$. Since, $F_n(x, y)$ and $F(x, y)$ are monotone in $x$ and $y$ both we have for $(x, y)$ in $D_{njk}$,

$$F_n(L_{nj}^{(1)}, L_{nk}^{(2)}) - F(L_{nj}^{(1)}, L_{nk}^{(2)}) \leq F_n(x, y) - F(x, y)$$

$$\leq F_n(L_{nj}^{(1)}, L_{nk}^{(2)}) - F(L_{nj}^{(1)}, L_{nk}^{(2)}). \quad (3.8)$$

For ease of notation, we may write $G_n(x, y) = G_n((x, y), \xi)$. Hence, for $G_n(x, y) = F_n(x, y) - F(x, y) - F_n(\xi) + F(\xi)$, we have for $(x, y) \in D_{njk}$

$$|G_n(x, y)| \leq \max_{(l, r) = (j, k) \text{ or } (j+1, k+1)} |G_n(L_{nl}^{(1)}, L_{nr}^{(2)})| + |\zeta_{njk}|, \quad (3.9)$$

where

$$\zeta_{njk} = F(L_{nj}^{(1)}, L_{nk}^{(2)}) - F(L_{nj}^{(1)}, L_{nk}^{(2)}). \quad (3.10)$$
Note that since for some \((x_0, y_0) \in D_{njk}\)
\[
\begin{align*}
\mathcal{P}_{V_n} &= \left( L_{n(j+1)}^{(1)} - L_{nj}^{(1)} \right) \frac{\partial F(x, y)}{\partial x} \bigg|_{(x, y) = (x_0, y_0)} + \left( L_{n(k+1)}^{(2)} - L_{nk}^{(2)} \right) \frac{\partial F(x, y)}{\partial y} \bigg|_{(x, y) = (x_0, y_0)} \\
&= O(a_n b_n^{-1}) \\
&= O(n^{-3/4}(\log n)^{(1+\eta)}),
\end{align*}
\]
(3.11)
where, we have chosen \(b_n \sim n^{1/4}\). Thus
\[
\sup_{(x, y) \in D_n} |G_n(x, y)| \leq \max_{|j| \leq b_n, |k| \leq b_n} |G_n(L_{nj}^{(1)}, L_{nk}^{(2)})| + O(n^{-3/4}(\log n)^{(1+\eta)}).
\]
(3.12)
In what follows we prove that for sufficiently large \(K\) and \(\varepsilon_n = Kn^{-3/4}(\log n)^{(1+\eta)}\), \(K > 0\),
\[
\sum_{n=1}^{\infty} \mathbb{P} \left( \max_{(\xi_1, \xi_2) \in [0,1] \times [0,1]} \sup_{(x, y) \in D_n(\xi)} |G_n((x, y), (\xi_1, \xi_2))| \geq \varepsilon_n \right) < \infty
\]
(3.13)
which will establish the lemma. Towards this end, consider \(L_{nj}^{(1)} > \xi_1\) and \(L_{nk}^{(2)} > \xi_2\), then we have
\[
nG_n(L_{nj}^{(1)}, L_{nk}^{(2)}) \sim \text{Bin}(n, p_{njk}) - np_{njk},
\]
(3.14)
where
\[
p_{njk} = F(L_{nj}^{(1)}, L_{nk}^{(2)}) - F(\xi)
\]
\[
= (L_{nj}^{(1)} - \xi_1) \frac{\partial F(x, y)}{\partial x} \bigg|_{(x, y) = (\tilde{x}, \tilde{y})} + (L_{nk}^{(2)} - \xi_2) \frac{\partial F(x, y)}{\partial y} \bigg|_{(x, y) = (\tilde{x}, \tilde{y})}.
\]
(3.15)
We can choose a constant \(c > 0\) such that \(p_{njk} < cn_n\) and then by using Bernstein’s inequality (see Eq. (12) of Bahadur, 1966) for sufficiently large \(n > N\)
\[
P[|G_n(L_{nj}^{(1)}, L_{nk}^{(2)})| \geq \varepsilon_n] \leq 2 \exp[-\delta_n],
\]
(3.16)
where
\[
\delta_n = \frac{c n_n^2}{2(c n_n + \varepsilon_n)}.
\]
(3.17)
Hence
\[
P \left[ \max_{|j| \leq b_n, |k| \leq b_n} \{ |G_n(L_{nj}^{(1)}, L_{nk}^{(2)})| \} \geq \varepsilon_n \right] \leq 8 b_n^2 \exp[-\delta_n] = \gamma_n, \text{ say.}
\]
(3.18)
We see that \(\lim_{n \to \infty} \log \gamma_n / \log n = \frac{1}{2} - K^2/2c\) which can be made much smaller than \(-2\), therefore for sufficiently large \(n > N\), \(\gamma_n \approx 1/n^u\) for \(u > 2\). Thus for sufficiently large \(n\), there exists \(u > 2\) such that,
\[
P \left[ \max_{|j| \leq b_n, |k| \leq b_n} |G_n(L_{nj}^{(1)}, L_{nk}^{(2)})| \geq \varepsilon_n \right] \leq \frac{1}{n^u}.
\]
(3.19)
Now, decomposing $[0,1] \times [0,1]$ into squares of length $1/a_n$, we get

$$P \left[ \sup_{(x,y) \in D_n(\xi)} |G_n(x,y)| \geq \varepsilon_n \right] \leq \frac{1}{a_n^2 n^u} \tag{3.20}$$

and then choosing $K$ large enough we find that

$$\sum_{n>N} P \left[ \sup_{(x,y) \in D_n(\xi)} |G_n(x,y)| \geq \varepsilon_n \right] < \sum_{n>N} \frac{1}{n^u^{-1}(\log n)^{2(1+\eta)}}. \tag{3.21}$$

Since, the right-hand side of the above equation converges for $u > 2$, the assertion in Eq. (3.13) follows.

**Proof of Theorem 3.2.** We may now write

$$\tilde{F}_n(x,y) - F_n(x,y) = \{\tilde{F}_n(x,y) - F^*_n(x,y)\} + \{F^*_n(x,y) - F_n(x,y)\}
= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}(\hat{\lambda}_{n1}, \hat{\lambda}_{n2}) \left\{ F_n\left( j, \frac{k}{\hat{\lambda}_{n1}}, \frac{k}{\hat{\lambda}_{n2}} \right) - F\left( j, \frac{k}{\hat{\lambda}_{n1}}, \frac{k}{\hat{\lambda}_{n2}} \right) \right. + \left. \{F^*_n(x,y) - F(x,y)\} \right\}
= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}(\hat{\lambda}_{n1}, \hat{\lambda}_{n2}) G_n((x,y), j, k)
+ \{F^*_n(x,y) - F(x,y)\}, \tag{3.22}$$

where $\tilde{\xi}_{jk} = (j/\hat{\lambda}_{n1}, k/\hat{\lambda}_{n2})$. Consider the set $N = \{0,1,2,\ldots\}$ and define for every $(x,y)$

$$N_{(x,y)} = \{(j,k): |j/\hat{\lambda}_{n1} - x| \leq (\hat{\lambda}_{n1}^{-1} (\log n)^{1+\eta})^{1/2}, |k/\hat{\lambda}_{n2} - y| \leq (\hat{\lambda}_{n2}^{-1} (\log n)^{1+\eta})^{1/2}\} \text{ and } N^c_{(x,y)} = (N \times N) \setminus N_{(x,y)}.$$

Then by Lemma 3.1 we find that

$$\sup_{(x,y) \in \mathbb{R}^2} \max_{(j,k) \in N_{(x,y)}} |G_n((x,y), \tilde{\xi}_{jk})| = O(n^{-3/4} (\log n)^{(1+\eta)}). \tag{3.23}$$

Further, by using Lemma 3.1 of Chaubey and Sen (1996), we have

$$\sum_{(x,y) \in N^c_{(x,y)}} w_{jk}(\hat{\lambda}_{n1}, \hat{\lambda}_{n2}) = o(n^{-1}). \tag{3.24}$$

Thus, combining Eqs. (3.23) and (3.24) we have

$$\sup_{(x,y) \in \mathbb{R}^2} |G_n((x,y), \tilde{\xi}_{jk})| = O(n^{-3/4} (\log n)^{(1+\eta)}). \tag{3.25}$$
Now, expanding $F\left(\frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}}\right)$ about $(x, y)$ we have

$$F\left(\frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}}\right) - F(x, y) = F'_x(x, y)(j/\hat{\lambda}_{1n} - x) + F'_y(x, y)(k/\hat{\lambda}_{2n} - y)$$

$$+ \frac{1}{2} F''_x(x, y)(j/\hat{\lambda}_{1n} - x)^2 + \frac{1}{2} F''_y(x, y)(k/\hat{\lambda}_{2n} - y)^2$$

$$+ F''_{xy}(x, y)(j/\hat{\lambda}_{1n} - x)(k/\hat{\lambda}_{2n} - y)$$

$$+ o \left( \left\| (x, y) - \left(\frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}}\right) \right\|^2 \right). \quad (3.26)$$

Using the properties of the Poisson probabilities as in Eq. (3.24) we have

$$\sum_{(x, y)\in N(x, y)} w_{jk}(\hat{\lambda}_{1n}x, \hat{\lambda}_{2n}y)(j/\hat{\lambda}_{1n} - x) = o(n^{-1}) \quad (3.27)$$

and

$$\sum_{(x, y)\in N(x, y)} w_{jk}(\hat{\lambda}_{1n}x, \hat{\lambda}_{2n}y)(k/\hat{\lambda}_{2n} - y) = o(n^{-1}). \quad (3.28)$$

Furthermore,

$$\sup\{ |j/\hat{\lambda}_{1n} - x|^2: (j, k) \in N(x, y) \} \leq \lambda_{1n}^{-1}(\log n)^{1+\eta}, \quad (3.29)$$

$$\sup\{ |k/\hat{\lambda}_{2n} - y|^2: (j, k) \in N(x, y) \} \leq \lambda_{2n}^{-1}(\log n)^{1+\eta} \quad (3.30)$$

and

$$\sup\{ |j/\hat{\lambda}_{1n} - x||k/\hat{\lambda}_{2n} - y|: (j, k) \in N(x, y) \} \leq \lambda_{1n}^{-1/2} \lambda_{2n}^{-1/2}(\log n)^{1+\eta}. \quad (3.31)$$

Combining Eqs. (3.26)–(3.31) we conclude that as $n \to \infty$,

$$|F'_n(x, y) - F(x, y)| = O(n^{-1}(\log n)^{1+\eta}). \quad (3.32)$$

The proof is completed by combining (3.22), (3.25) and (3.32). □

**Remark 3.1.** The above theorems hold for $\tilde{S}_n(x)$ by basically mimicking the same proof.

**Remark 3.2.** The same order as given in the previous theorem is also achieved using the truncated Poisson weights.

**Remark 3.3.** As a consequence of the above theorem the asymptotic distribution of the smooth estimator is the same as that of the empirical distribution function.
3.2. Properties of the smooth estimator of the density function

The next theorem establishes the strong uniform consistency of the multivariate density estimator.

**Theorem 3.3.** Assuming that the density function \( f \) is bounded with bounded first derivatives and \( \lambda_{mn} = O(n^{2/d}) \) for some \( \alpha < \frac{3}{4} \), for \( i = 1, 2, \ldots, d \), we have

\[
\sup_{x \in \mathbb{R}^d} |\hat{f}_n(x) - f(x)| \to 0 \text{ a.s. as } n \to \infty. \tag{3.33}
\]

**Proof.** We provide the proof below for the bivariate case only. The general case for \( d > 2 \) can be handled similarly. First note that by the fundamental theorem of calculus, for any fixed \( x \), \( f(x, y) = \partial / \partial x G(x, y) \) where \( G(x, y) = \int_0^y f(x, v) \, dv \) is monotone and bounded in \( y \). Hence, for any \( \varepsilon > 0 \), there exists constant \( c = c_\varepsilon \), such that

\[
G(x, y + z) - G(x, y) < \varepsilon \text{ for } z \geq 0 \text{ and } \forall y \geq c.
\]

Since the left-hand side equals \( \int_y^{y+z} f(x, v) \, dv \), using the mean value theorem, we conclude by choosing \( z \) sufficiently large, so that \( \varepsilon / z \) is small, that for a fixed \( x \)

\[
f(x, y) \leq \varepsilon' \text{ for every } y \geq c.
\]

A similar argument gives that, as \( n \to \infty \),

\[
\hat{f}_n(x, y) \leq \varepsilon' \text{ a.s. for every } y \geq c.
\]

Using, the same analysis for a fixed \( y \), we have for some \( c > 0 \), \( d > 0 \), \( \sup|\hat{f}_n(x, y) - f(x, y)| : x \geq c, y \geq d | < 2\varepsilon' \), a.s. as \( n \to \infty \). Thus, for proving the theorem we only need to establish

\[
\sup_{(x, y) \in [0, c] \times [0, d]} |\hat{f}_n(x, y) - f(x, y)| \to 0 \text{ a.s. as } n \to \infty. \tag{3.34}
\]

For this purpose we decompose the estimator in Eq. (2.8) into two terms as follows:

\[
\hat{f}_n(x, y) = T_{1n}(x, y) + T_{2n}(x, y), \tag{3.35}
\]

where

\[
T_{1n}(x, y) = \lambda_{1n} \lambda_{2n} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}(\lambda_{1n} x, \lambda_{2n} y) [P_{njk} - P_{jk}],
\]

\[
T_{2n}(x, y) = \lambda_{1n} \lambda_{2n} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}(\lambda_{1n} x, \lambda_{2n} y) P_{jk} \tag{3.36}
\]
with $P_{njk}$ and $P_{jk}$ defined as

\[
P_{njk} = \left[ S_n \left( \frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) - S_n \left( \frac{j+1}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) - S_n \left( \frac{j}{\hat{\lambda}_{1n}}, \frac{k+1}{\hat{\lambda}_{2n}} \right) \right. \\
+ S_n \left( \frac{j+1}{\hat{\lambda}_{1n}}, \frac{k+1}{\hat{\lambda}_{2n}} \right) - S \left( \frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) + \left. S \left( \frac{j+1}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) \right] ,
\]

\[
P_{jk} = \left[ S \left( \frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) - S \left( \frac{j+1}{\hat{\lambda}_{1n}}, \frac{k+1}{\hat{\lambda}_{2n}} \right) \right. \\
- \left. S \left( \frac{j+1}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) + S \left( \frac{j+1}{\hat{\lambda}_{1n}}, \frac{k+1}{\hat{\lambda}_{2n}} \right) \right] . \tag{3.37}
\]

By following similar steps as in the proof of Lemma 3.1, we find that

\[
\sup_{(x,y) \in \mathbb{R}^2_+} |T_{1n}(x, y)| = O(n^{2/3}(\log n)^{1+\eta}) \text{ a.s. as } n \to \infty \tag{3.38}
\]

which converges to 0, a.s. as $n \to \infty$ for $\alpha < \frac{3}{4}$. For the analysis of the term $T_{2n}(x, y)$ we first note that

\[
P_{jk} = \int_{k+1/\hat{\lambda}_{2n}}^{j+1/\hat{\lambda}_{1n}} \int_{j/\hat{\lambda}_{1n}}^{k/\hat{\lambda}_{2n}} f(u, v) \, du \, dv. \tag{3.39}
\]

Next, using the bivariate Taylor’s expansion of $f(u, v)$ for $(u, v) \in [j/\hat{\lambda}_{1n}, j+1/\hat{\lambda}_{1n}] \times [k/\hat{\lambda}_{2n}, k+1/\hat{\lambda}_{2n}]$ we find that

\[
f(u, v) = f \left( \frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) + \left( u - \frac{j}{\hat{\lambda}_{1n}} \right) f'_x \left( \frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} + h \left( \frac{u - j}{\hat{\lambda}_{1n}} \right), \frac{k}{\hat{\lambda}_{2n}} + h \left( \frac{v - k}{\hat{\lambda}_{2n}} \right) \right) \\
+ \left( u - \frac{j}{\hat{\lambda}_{1n}} \right) f'_y \left( \frac{j}{\hat{\lambda}_{1n}} + h \left( \frac{u - j}{\hat{\lambda}_{1n}} \right), \frac{k}{\hat{\lambda}_{2n}} + h \left( \frac{v - k}{\hat{\lambda}_{2n}} \right) \right) \tag{3.40}
\]

for some $0 < |h| < 1$, where $f'_x$ denotes the derivative of $f$ with respect to $x$, and therefore

\[
P_{jk} = \frac{1}{\hat{\lambda}_{1n}\hat{\lambda}_{2n}} f \left( \frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) + O \left( \frac{1}{\hat{\lambda}_{1n}\hat{\lambda}_{2n}} \left( \frac{1}{\hat{\lambda}_{1n}} \wedge \frac{1}{\hat{\lambda}_{2n}} \right) \right) . \tag{3.41}
\]

Thus, we have from Eq. (3.36)

\[
T_{2n}(x, y) = \sum_{j,k} w_{jk}(\lambda_{1n} x, \lambda_{2n} y) f \left( \frac{j}{\hat{\lambda}_{1n}}, \frac{k}{\hat{\lambda}_{2n}} \right) + O \left( \frac{1}{\hat{\lambda}_{1n}} \wedge \frac{1}{\hat{\lambda}_{2n}} \right) . \tag{3.42}
\]
Thus

\[
\sup_{(x,y) \in C} |T_{2n}(x, y) - f(x, y)| \leq \sup_{(x,y) \in C} \left| \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}(\lambda_{1n} x, \lambda_{2n} y) \right| + O \left( \frac{1}{\lambda_{1n}} \land \frac{1}{\lambda_{2n}} \right).
\]

(3.43)

The first term on the right of Eq. (3.43) tends to zero, by virtue of Lemma 2.1 and the second term tends to zero by the assumption on \(\lambda_{in}\). This completes the proof. \(\Box\)

**Remark 3.4.** The existence and the boundedness of first derivatives are not necessary for the validity of the above theorem. We may assume a weaker Lipschitz condition on \(f(x)\), namely, assume that for some \(\gamma > 0\)

\[
|f(x) - f(u)| \leq K ||(x - u)||^\gamma \quad \text{for every } x, u \in \mathbb{R}^d,
\]

where, \(K < \infty\) does not depend on \(x, u\). In this case, the second term in (3.42) is bounded by \(2K \max(\lambda_{in}^{-\gamma}, i = 1, \ldots, d)\) and the proof goes through.

Now we establish the asymptotic normality of the multivariate density estimator as given in the following theorem.

**Theorem 3.4.** Consider a non-stochastic sequence \(\{\lambda_{in}\}_{n=1}^{\infty}, i = 1, 2, \ldots, d\) such that \(\lambda_{in} = O(n^{2/(4+d)})\). Further assume that the first order partial derivatives are Lipschitz continuous, then for every fixed \(x \in \mathbb{R}^d\),

\[
n^{2(4+d)}(f_n(x) - f(x)) - \left( \frac{1}{2} \right) b(x) \xrightarrow{D} N(0, \sigma^2(x)) \quad \text{as } n \to \infty,
\]

(3.44)

where

\[
b(x) = \lim_{n \to \infty} n^{2(4+d)} \sum_{i=1}^{d} \lambda_{in}^{-1} f'_{x_i}(x)
\]

(3.45)

and

\[
\sigma^2(x) = \lim_{n \to \infty} \frac{1}{2^{d/2} n^{d/2} \prod \lambda_{in}^{1/2} \prod \lambda_{in}^{1/2} f(x)}.
\]

(3.46)

**Proof.** For making it transparent we outline the proof below for the bivariate case. The general case follows similarly. Note that in light of the Lipschitz continuity of the first derivatives of \(f(x, y)\) we may refine the representation of \(T_{2n}(x, y)\) given in Eq. (3.42) as

\[
T_{2n}(x, y) = \sum_{j,k} w_{jk}(\lambda_{1n} x, \lambda_{2n} y) f \left( \frac{j}{\lambda_{1n}}, \frac{k}{\lambda_{2n}} \right)
\]

\[
+ \frac{1}{\lambda_{1n}} \sum_{j,k} w_{jk}(\lambda_{1n} x, \lambda_{2n} y) f'_{x} \left( \frac{j}{\lambda_{1n}}, \frac{k}{\lambda_{2n}} \right).
\]
\[ + \frac{1}{\lambda_{2n}} \sum_{j,k} w_{jk}(\lambda_{1n} x, \lambda_{2n} y) f_{y}' \left( \frac{j}{\lambda_{1n}}, \frac{k}{\lambda_{2n}} \right) \]
\[ + \mathcal{O} \left[ \left\{ \frac{1}{\lambda_{1n}} \left( \frac{\lambda_{2}^2}{\lambda_{1n}^2} + \frac{\lambda_{2}^2}{\lambda_{2n}^2} \right)^{\delta_1/2} \right\} \wedge \left\{ \frac{1}{\lambda_{2n}} \left( \frac{\lambda_{1}^2}{\lambda_{1n}^2} + \frac{\lambda_{1}^2}{\lambda_{2n}^2} \right)^{\delta_2/2} \right\} \right] , \quad (3.47) \]

where \( \delta_1 \) and \( \delta_2 \) are positive constants such that
\[
|f_x'(x, y) - f_x'(u, v)| \leq K_1 [(x - u)^2 + (y - v)^2]^{\delta_1/2} \]
and
\[
|f_y'(x, y) - f_y'(u, v)| \leq K_2 [(x - u)^2 + (y - v)^2]^{\delta_2/2} .
\]

Thus, we have the following almost sure representation of \( \tilde{f}_n(x, y) \), using Eqs. (3.35), (3.38) and (3.47),
\[
\tilde{f}_n(x, y) = f(x, y) + \frac{1}{\lambda_{1n}} f_x'(x, y) + \frac{1}{\lambda_{2n}} f_y'(x, y)
\]
\[ + \mathcal{O} \left[ \left\{ \frac{1}{\lambda_{1n}} \left( \frac{\lambda_{2}^2}{\lambda_{1n}^2} + \frac{\lambda_{2}^2}{\lambda_{2n}^2} \right)^{\delta_1/2} \right\} \wedge \left\{ \frac{1}{\lambda_{2n}} \left( \frac{\lambda_{1}^2}{\lambda_{1n}^2} + \frac{\lambda_{1}^2}{\lambda_{2n}^2} \right)^{\delta_2/2} \right\} \right] . \quad (3.48) \]

We will show that \( \text{Var}(T_{1n}(x, y)) = \mathcal{O}(\prod \lambda_{in}^{1/2} / n) \), hence the MSE of \( \tilde{f}_n(x, y) \) is \( \mathcal{O}(\prod \lambda_{in}^{1/2} / n) + \mathcal{O}(1/\lambda_{1n}^2 \wedge 1/\lambda_{2n}^2) \). Thus, we need to limit \( \lambda_{in} \), bounded below by \( cn^{2/(4+d)} \) for some \( c > 0 \). Choosing, \( \lambda_{in} = o(n^{2/(4+d)}) \), enhances the bias term to \( \mathcal{O}(n^{-2(4+d)}) \), however, for, \( \lambda_{in} \sim n^{2/(4+d)} \), the normalizing factor, \( n^{2/(4+d)} \), of the variance and the bias terms are of the same order. Therefore, to complete the proof of the theorem, we need the asymptotic normality of \( T_{1n}(x, y) \). For this purpose, we may write,
\[
T_{1n}(x, y) = \lambda_{1n} \lambda_{2n} \sum_{jk} w_{jk}(\lambda_{1n} x, \lambda_{2n} y) Z_{nkj},
\]
where \( Z_{nkj} \) represent the multinomial proportions with probabilities \( P_{jk} \). As such we have,
\[
\text{Var}(T_{1n}(x, y)) = \frac{\lambda_{1n}^2 \lambda_{2n}^2}{n} \left[ \sum_{jk} w_{jk}^2 P_{jk} - \left( \sum_{jk} w_{jk} P_{jk} \right)^2 \right] , \quad (3.49)
\]
where, we have abbreviated \( w_{jk}(\lambda_{1n} x, \lambda_{2n} y) = w_{jk} \). First, we find basically, following similar steps as in the proof of Theorem 3.3, that
\[
\frac{\lambda_{1n}^2 \lambda_{2n}^2}{n} \left( \sum_{jk} w_{jk} P_{jk} \right)^2 = \frac{1}{n} f^2(x, y) + \mathcal{O} \left( \left( \frac{1}{\lambda_{1n}^2} \wedge \frac{1}{\lambda_{2n}^2} \right) \frac{1}{n} \right) ,
\]
which is of the order \( o(n^{4/(4+d)}) \), hence, we need to work out the approximation for \( \sum_{jk} w_{jk}^2 P_{jk} \). Towards this end we work with the leading term of the sum given by
(see Eq. 3.41) \( \frac{\lambda_1 \lambda_2}{n} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} f(j/\lambda_1, k/\lambda_2) w_{jk}^2 (\lambda_1 x, \lambda_2 y) \). For a fixed \( j \), using the Stirling’s formula to the factorials and approximating the sum by integral we have,

\[
\frac{\lambda_2}{n} \sum_{k=0}^{\infty} f\left( \frac{j}{\lambda_1}, \frac{k}{\lambda_2} \right) w_{jk}^2 (\lambda_1 x, \lambda_2 y)
\]

\[
= \frac{\lambda_2}{n} \sum_{k=0}^{\infty} f\left( \frac{j}{\lambda_1}, \frac{k}{\lambda_2} \right) w_{jk}^2 (\lambda_2 y)
\]

\[
\simeq \frac{\lambda_2}{2\pi n} \int_{-\infty}^{\infty} e^{-2t^2} f\left( \frac{j}{\lambda_1}, y + t \sqrt{y/\lambda_2} \right) \left( t + \sqrt{\lambda_2 y} \right)^{-1} dt
\]

\[
\simeq \frac{1}{2} \sqrt{\frac{1}{\pi y}} f\left( \frac{j}{\lambda_1}, y \right) \frac{\sqrt{\lambda_2}}{n}.
\]

The procedure now allows us to obtain, by summing over \( j \),

\[
\frac{\lambda_1 \lambda_2}{n} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{jk}^2 (\lambda_1 x, \lambda_2 y) f\left( \frac{j}{\lambda_1}, \frac{k}{\lambda_2} \right) \simeq \frac{1}{2^d} \frac{1}{\pi^{d/2}} f(x, y) \frac{\sqrt{\lambda_1 \lambda_2}}{n}.
\]

Following the same procedure for the general case gives the result

\[
\text{Var}(\tilde{f}_n(x)) = \frac{1}{2^d} \frac{1}{\pi^{d/2}} \frac{1}{\prod x_j^{1/2}} f(x) \prod \lambda_i^{1/2} \frac{1}{n}
\]

which completes the proof of the theorem. □

Acknowledgements

The research of the first author (YPC) is partially supported from a research grant from NSERC of Canada. The authors would like to thank the referees for some constructive comments on the first draft of the paper.

References


