Journal of Nonparametric Statistics
Vol. 20, No. 8, November 2008, 661–677

Kernel estimation of multivariate cumulative distribution function

Rong Liu and Lijian Yang*

Department of Statistics & Probability, Michigan State University, East Lansing, MI, USA

(*Received 26 December 2007; final version received 5 July 2008*)

A smooth kernel estimator is proposed for multivariate cumulative distribution functions (cdf), extending the work of Yamato [H. Yamato, *Uniform convergence of an estimator of a distribution function*, Bull. Math. Statist. 15 (1973), pp. 69–78.] on univariate distribution function estimation. Under assumptions of strict stationarity and geometrically strong mixing, we establish that the proposed estimator follows the same pointwise asymptotically normal distribution of the empirical cdf, while the new estimator is a smooth instead of a step function as the empirical cdf. We also show that under stronger assumptions the smooth kernel estimator has asymptotically smaller mean integrated squared error than the empirical cdf, and converges to the true cdf uniformly almost surely at a rate of $(n^{-1/2} \log n)$. Simulated examples are provided to illustrate the theoretical properties. Using the smooth estimator, survival curves for US gross domestic product (GDP) growth are estimated conditional on the unemployment growth rate to examine how GDP growth rate depends on the unemployment policy. Another example of gold and silver price returns is given.

Keywords: bandwidth; Berry–Esseen bound; GDP; gold price return; kernel; mean integrated squared error; rate of convergence; silver price return; strongly mixing; survival function; unemployment rate

1. Introduction

The estimation of probability density functions (pdf) and cumulative distribution functions (cdf) occupies a central place in applied data analysis in the social sciences. While many statisticians and econometricians are familiar with various smooth nonparametric estimators of pdf, the smooth estimation of cdf has not been investigated as much, see Li and Racine [1], Sections 1.4 and 1.5. To properly define the problem, let $\{\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^{\mathrm{T}}\}_{i=1}^{n}$ be a geometrically α-mixing and strictly stationary sequence of d-dimensional variables, with a common pdf $f \in C^{(p+1)}(R^d)$ and cdf $F \in C^{(p+d+1)}(R^d)$, in which p is an odd integer. Traditionally, F is estimated by the empirical cdf $\widehat{F}(\mathbf{x}) = n^{-1} \sum_{i=1}^{n} I\{\mathbf{X}_i \leq \mathbf{x}\}$, whose theoretical properties have been well known. One obvious drawback of $\widehat{F}$ is that it is a step function even when the true cdf F is smooth.

*Corresponding author. Email: yang@stt.msu.edu

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485250802326391
http://www.informaworld.com

Yamato [2] proposed a smooth estimator of F by integrating a kernel density estimator of the density f. To be precise, define the following kernel estimator of F

$$\hat{F}(\mathbf{x}) = \hat{F}_n(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} \hat{f}(\mathbf{u})d\mathbf{u} = n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}, \quad \forall \mathbf{x} \in R^d \tag{1}$$

where $\hat{f}(\mathbf{u})$ is the standard d-dimensional kernel density estimator (kde) of $f(\mathbf{u})$ (see [3])

$$\hat{f}(\mathbf{u}) = n^{-1}\sum_{i=1}^{n} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u}), \quad K_{\mathbf{h}}(\mathbf{u}) = \prod_{\alpha=1}^{d}\frac{1}{h_\alpha}K\left(\frac{u_\alpha}{h_\alpha}\right), \quad \mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$$

in which $\mathbf{h} = (h_1, \ldots, h_d)^{\mathrm{T}}$ are positive numbers depending on the sample size n, called bandwidths.

Theoretical properties of $\hat{F}(\mathbf{x})$ as an estimator of the unknown true distribution function $F(\mathbf{x})$ have been investigated by several authors for the case of $d = 1$ and under i.i.d assumptions, see *e.g.* Yamato [2], Reiss [4], Falk [5] and more recently Cheng and Peng [6]. For feasible econometric applications of univariate kernel estimation of cdf, such as to the testing of stochastic dominance, see Li and Racine [1, p. 23] and the references therein.

In this paper, we examine under strong mixing assumption and for arbitrary dimension d, the local property of $\hat{F}(\mathbf{x})$ in terms of pointwise asymptotic distribution and its global property in terms of mean integrated squared error (MISE) and maximal deviation. We have (1) proven that the smooth estimator $\hat{F}(\mathbf{x})$ behaves asymptotically similar to the empirical cdf $\widehat{F}(\mathbf{x})$ at any point $\mathbf{x}$, (2) obtained its asymptotic mean integrated squared error (AMISE) and (3) established its uniform almost sure convergence rate.

The paper is organised as follows. In Section 2, we give Theorems 1, 2 and 3, the main results on pointwise, MISE and uniform asymptotics. In Section 3, we describe a data-driven rule to select the asymptotically optimal bandwidth vector $\mathbf{h}$, which makes the MISE of $\hat{F}$ asymptotically smaller than that of the empirical cdf $\widehat{F}$ according to Theorem 5, another compelling reason that $\hat{F}$ is preferable over $\widehat{F}$ other than smoothness. In Section 4, we present Monte Carlo evidence that corroborates with the theory and illustrates the use of $\hat{F}$ with two real data examples. The first real data example illustrates the stochastic dependence of gross domestic product (GDP) growth rate on the unemployment growth rate in the US economy. The second example shows that gold and silver are substitute goods and their prices are strongly associated. All technical proofs are in the Appendix.

2. Asymptotic results

Throughout this paper, we denote

$$h_{\max} = \max(h_1, \ldots, h_d), \quad h_{\mathrm{prod}} = h_1 \times \cdots \times h_d$$

and for any $x \in R$, $\tilde{K}(x) = \int_{-\infty}^{x} K(u)du$, and $\tilde{K}(\mathbf{x}) = \prod_{\alpha=1}^{d}\tilde{K}(x_\alpha)$ for any vector $\mathbf{x} = (x_1, \ldots, x_d)^{\mathrm{T}}$. Then $\tilde{K}(\mathbf{x}) \equiv 0$ unless $\mathbf{x} \geq -1$ and $\tilde{K}(\mathbf{x}) \equiv 1$ if $\mathbf{x} \geq 1$, where for any two vectors $\mathbf{x} = (x_1, \ldots, x_d)^{\mathrm{T}}$, $\mathbf{y} = (y_1, \ldots, y_d)^{\mathrm{T}}$, $\mathbf{x} \geq \mathbf{y}$ if and only if $x_\alpha \geq y_\alpha, \forall \alpha = 1, \ldots, d$. It is easily verified that $\int_{-1}^{1}\tilde{K}(w)dw = 1$. We also denote $\mu_{p+1}(K) = \int_{-1}^{1} K(u)u^{p+1}du$, $D(K) = 1 - \int_{-1}^{1}\tilde{K}^2(w)dw$. For any vector $\mathbf{x} = (x_1, \ldots, x_d)^{\mathrm{T}}$ and $\forall \alpha = 1, \ldots, d$, we denote $\mathbf{x}_{\_\alpha} = (x_1, \ldots, x_{\alpha-1}, x_{\alpha+1}, \ldots, x_d)^{\mathrm{T}}$ and with slight abuse of notation, write $\mathbf{x} = (x_\alpha, \mathbf{x}_{\_\alpha})^{\mathrm{T}}$.

We next list some basic assumptions.

(A1) *The cumulative distribution function* $F \in C^{(p+d+1)}(R^d)$, *in which p is an odd integer, while all* $(p+d+1)$*-th partial derivatives of F belong to* $L_1(R^d)$ *and* $\max_{\mathbf{x}\in R^d} |f(\mathbf{x})| \le C$.

(A2) *There exist positive constants* K_0 *and* λ_0 *such that* $\alpha(k) \le K_0 \exp(-\lambda_0 k)$ *holds for all k, where the kth order strong mixing coefficient of the strictly stationary process* $\{\mathbf{X}_s\}_{s=-\infty}^{\infty}$ *is defined as*

$$\alpha(k) = \sup_{B\in\sigma\{\mathbf{X}_s, s\le t\}, C\in\sigma\{\mathbf{X}_s, s\ge t+k\}} |P(B\cap C) - P(B)P(C)|, \quad k \ge 1.$$

(A3) *As* $n \to \infty$, $nh_{\mathrm{prod}} \to \infty$, $n^{1/2}h_{\mathrm{prod}}/(\log n)^{1/2} + n^{1/2}h_{\max}^{p+1} \to 0$.

(A4) *The univariate kernel function* $K(\cdot)$ *is of* $(p+1)$*-th order, supported on* $[-1, 1]$, *Lipschitz continuous.*

Assumptions (A1)–(A4) are all typical conditions in time series smoothing literature, see Bosq [7, Chap. 2] for similar or even stronger assumptions. Elementary arguments show that $D(K) > 0$ under Assumption (A4).

The following theorem concerns the asymptotic distribution of $\hat{F}$ given in Equation (1) at any $\mathbf{x} \in R^d$.

THEOREM 1 *Under Assumptions* (A1)–(A4), $\forall x \in R^d$ *as* $n \to \infty$

$$\sqrt{nV^{-1}(\mathbf{x})}(\hat{F}(\mathbf{x}) - F(\mathbf{x})) \longrightarrow_d N(0,1),$$

where

$$V(\mathbf{x}) = \sum_{l=-\infty}^{\infty} \gamma(l), \quad \gamma(l) = EI\{\mathbf{X}_i \le \mathbf{x}\}I\{\mathbf{X}_{i+l} \le \mathbf{x}\} - F^2(\mathbf{x}).$$

Theorem 1 shows that the smooth estimator $\hat{F}(\mathbf{x})$ has asymptotically the same distribution as the empirical cdf $\widehat{F}(\mathbf{x})$. In particular, for i.i.d. process $\{\mathbf{X}_s\}_{s=-\infty}^{\infty}$, the asymptotic variance function $V(\mathbf{x})$ reduces to the more familiar form of $\gamma(0) = F(\mathbf{x})\{1 - F(\mathbf{x})\}$.

The global performance of $\hat{F}(\mathbf{x})$ as an estimator of $F(\mathbf{x})$ can be measured in terms of MISE and maximal deviation

$$\mathrm{MISE}(\hat{F}) = \mathrm{MISE}(\hat{F}; \mathbf{h}) = E\int \{\hat{F}(\mathbf{x}) - F(\mathbf{x})\}^2 dF(\mathbf{x}), \tag{2}$$

$$D_n(\hat{F}) = D_n(\hat{F}; \mathbf{h}) = \sup_{\mathbf{x}\in R^d} |\hat{F}(\mathbf{x}) - F(\mathbf{x})|. \tag{3}$$

The next two theorems give an asymptotic formula of $\mathrm{MISE}(\hat{F})$ and an almost sure rate of $D_n(\hat{F})$.

THEOREM 2 *Under Assumptions* (A1)–(A4), *as* $n \to \infty$,

$$\mathrm{MISE}(\hat{F}; \mathbf{h}) = \mathrm{AMISE}(\hat{F}; \mathbf{h}) + o(h_{\max}^{2p+2} + n^{-1}h_{\max})$$

in which the AMISE is

$$\mathrm{AMISE}(\hat{F}; \mathbf{h}) = \frac{\int V(\mathbf{x})dF(\mathbf{x})}{n} + \frac{\mu_{p+1}^2(K)}{(p+1)!^2}\sum_{\alpha,\beta=1}^{d} h_\alpha^{p+1}h_\beta^{p+1}B_{\alpha\beta,p+1}(F) - \frac{D(K)\sum_{\alpha=1}^{d} h_\alpha C_\alpha(F)}{n}$$

with

$$B_{\alpha\beta,p+1}(F)=\int\frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}}\frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\beta^{p+1}}dF(\mathbf{x}),\quad C_\alpha(F)=\int\frac{\partial F(\mathbf{x})}{\partial x_\alpha}dF(\mathbf{x}),\ \forall\alpha,\beta=1,\ldots,d.$$

THEOREM 3 *Under Assumptions* (A1)–(A4)*, as* $n\to\infty$, $D_n(\hat{F})=O_{a.s.}(n^{-1/2}\log n)$ *while for i.i.d.* $\mathbf{X}_1,\ldots,\mathbf{X}_n$, $D_n(\hat{F})=O_{a.s.}(n^{-1/2}(\log n)^{1/2})$.

The first term $n^{-1}\int V(\mathbf{x})dF(\mathbf{x})$ in the formula of AMISE$(\hat{F};\mathbf{h})$ is the exact MISE of the empirical cdf $\widehat{F}$. We are unaware of any published results on the MISE or the strong uniform rate of convergence for smooth estimation of multivariate distribution function based on strongly mixing data, as in Theorems 2 and 3. Since Assumptions (A1)–(A4) are mild, these strong theoretical results hold for most multiple time series data with continuous distributions.

In the next section we describe how Theorem 2 is used to compute a data-driven bandwidth vector for implementing the smoothed estimator $\hat{F}$.

3. Bandwidth selection

To have insight into the minimisation of AMISE$(\hat{F};\mathbf{h})$ given in Theorem 2, define a function $Q: R_+^d\times M_+(d)\times R_+^d$ for elementwise positive vectors $\mathbf{v}=(v_1,\ldots,v_d)^{\mathrm{T}}$, $\mathbf{a}=(a_1,\ldots,a_d)^{\mathrm{T}}\in R_+^d=(0,+\infty)^d$ and $\mathbf{M}=(M_{\alpha\beta})_{\alpha,\beta=1}^d\in M_+(d)$, the set of all positive definite $d\times d$ matrices:

$$Q(\mathbf{v},\mathbf{M},\mathbf{a})=\sum_{\alpha,\beta=1}^d v_\alpha v_\beta M_{\alpha\beta}-\sum_{\alpha=1}^d a_\alpha v_\alpha^{1/(p+1)}=\mathbf{v}^T\mathbf{M}\mathbf{v}-\mathbf{a}^T\mathbf{v}^{1/(p+1)}$$

in which $\mathbf{v}^{1/(p+1)}=(v_1^{1/(p+1)},\ldots,v_d^{1/(p+1)})^{\mathrm{T}}$. In the following, we denote for any d-dimensional vector $\mathbf{a}=(a_1,\ldots,a_d)^{\mathrm{T}}$, the $d\times d$ diagonal matrix whose $(\alpha\alpha)$-th element is a_α, $\alpha=1,\ldots,d$ as diag$(\mathbf{a})$. The following theorem is easily proved similar to Yang and Tschernig [8].

THEOREM 4 (*i*) *The gradient and Hessian matrices of* $Q(\mathbf{v},\mathbf{M},\mathbf{a})$ *with respect to* $\mathbf{v}$ *are*

$$\frac{\partial}{\partial\mathbf{v}}Q(\mathbf{v},\mathbf{M},\mathbf{a})=\{\mathrm{diag}(M_{\alpha\alpha})_{\alpha=1}^d+\mathbf{M}\}\mathbf{v}-\frac{1}{p+1}\mathrm{diag}(\mathbf{a})\mathbf{v}^{1/(p+1)-1},$$

$$\frac{\partial^2}{\partial\mathbf{v}\partial\mathbf{v}^T}Q(\mathbf{v},\mathbf{M},\mathbf{a})=\mathrm{diag}(M_{\alpha\alpha})_{\alpha=1}^d+\mathbf{M}+\frac{p}{(p+1)^2}\mathrm{diag}\big(a_\alpha v_\alpha^{1/(p+1)-2}\big)_{\alpha=1}^d$$

the Hessian matrix of $Q(\mathbf{v},\mathbf{M},\mathbf{a})$ *is positive definite, hence the function* $Q(\mathbf{v},\mathbf{M},\mathbf{a})$ *is strictly convex in* $\mathbf{v}$. *(ii) For any* $\mathbf{a}\in R_+^d$, $\mathbf{M}\in M_+(d)$*, there exists a unique* $\mathbf{v}\in R_+^d$ *which minimises* $Q(\mathbf{v},\mathbf{M},\mathbf{a})$*, denoted as* $\mathbf{v}(\mathbf{M},\mathbf{a})$*, which satisfies* $(\partial/\partial\mathbf{v})Q(\mathbf{v},\mathbf{M},\mathbf{a})=\mathbf{0}$. *In addition,* $Q\{\mathbf{v}(\mathbf{M},\mathbf{a}),\mathbf{M},\mathbf{a}\}<0$ *for any* $\mathbf{a}\in R_+^d$, $\mathbf{M}\in M_+(d)$. *(iii) Finally, for any* $c_{\mathbf{M}}, c_{\mathbf{a}}>0$

$$Q\big(c_{\mathbf{a}}^{(p+1)/(2p+1)}c_{\mathbf{M}}^{-(p+1)/(2p+1)}\mathbf{v},\mathbf{c_M}M,\mathrm{c}_{\mathbf{a}}\mathbf{a}\big)=c_{\mathbf{a}}^{(2p+2)/(2p+1)}c_{\mathbf{M}}^{-1/(2p+1)}Q(\mathbf{v},\mathbf{M},\mathbf{a}),$$

$$\mathbf{v}(\mathbf{c_M}M,c_{\mathbf{a}}\mathbf{a})=c_{\mathbf{a}}^{(p+1)/(2p+1)}c_{\mathbf{M}}^{-(p+1)/(2p+1)}\mathbf{v}(\mathbf{M},\mathbf{a}).$$

To make use of Theorem 4, we make an additional assumption on F,

(A5) *The matrices* $\mathbf{B}_{p+1}(F)=\{B_{\alpha\beta,p+1}(F)\}_{\alpha,\beta=1}^d\in M_+(d)$ *and* $\mathbf{C}(F)=\{C_\alpha(F)\}_{\alpha=1}^d\in R_+^d$.

Theorem 2, Theorem 4(ii) and Assumption (A5) ensure the existence of a unique optimal bandwidth vector $\mathbf{h}_{\mathrm{opt}}$ that minimises

$$\mathrm{AMISE}(\hat{F};\mathbf{h}) = \frac{\int V(\mathbf{x})dF(\mathbf{x})}{n} + Q\left(\mathbf{h}^{p+1}, \frac{\mu_{p+1}^2(K)}{(p+1)!^2}\mathbf{B}_{p+1}(F), n^{-1}D(K)\mathbf{C}(F)\right).$$

Theorem 4(iii) then implies that

$$\begin{aligned}\mathbf{h}_{\mathrm{opt}} = \mathbf{h}_{\mathrm{opt}}(n,K,F) &= \mathbf{v}^{1/(p+1)}\left(\frac{\mu_{p+1}^2(K)}{(p+1)!^2}\mathbf{B}_{p+1}(F), n^{-1}D(K)\mathbf{C}(F)\right)\\ &= n^{-1/(2p+1)}\left\{\frac{\mu_{p+1}^2(K)}{D(K)(p+1)!^2}\right\}^{-1/(2p+1)}\mathbf{v}^{1/(p+1)}(\mathbf{B}_{p+1}(F),\mathbf{C}(F)).\end{aligned}$$

Thus to obtain the optimal bandwidth vector $\mathbf{h}_{\mathrm{opt}}$, one computes exactly the factors involving n and K in the previous expression, and estimate the following factor

$$\theta = \theta(F) = (\theta_1,\ldots,\theta_d)^{\mathrm{T}} = (\theta_1(F),\ldots,\theta_d(F))^{\mathrm{T}} = \mathbf{v}^{1/(p+1)}(\mathbf{B}_{p+1}(F),\mathbf{C}(F)).$$

The next theorem follows from the negativity result in Theorem 4(ii).

THEOREM 5 *Under Assumptions* (A1)–(A5)*, $\hat{F}$ has asymptotically smaller MISE than the empirical cdf $\widehat{F}$. Specifically,* $\mathrm{MISE}(\widehat{F}) = n^{-1}\int V(\mathbf{x})dF(\mathbf{x})$ *and as $n\to\infty$*

$$\begin{aligned}\mathrm{MISE}(\hat{F};\mathbf{h}_{\mathrm{opt}}) &= \mathrm{MISE}(\widehat{F}) + n^{-(2p+2)/(2p+1)}C(K,F) + o(n^{-(2p+2)/(2p+1)}),\\ C(K,F) &= \left\{\frac{D(K)^{2p}\mu_{p+1}^2(K)}{(p+1)!^2}\right\}^{-1/(2p+1)} Q(\mathbf{v}(\mathbf{B}_{p+1}(F),\mathbf{C}(F)),\mathbf{B}_{p+1}(F),\mathbf{C}(F)) < 0.\end{aligned}$$

Following Yang and Tschernig [8], we define a plug-in asymptotic optimal bandwidth vector

$$\hat{\mathbf{h}}_{\mathrm{opt}} = \left\{\frac{n\mu_{p+1}^2(K)}{C(K)(p+1)!^2}\right\}^{-1/(2p+1)}\mathbf{v}^{1/(p+1)}(\hat{\mathbf{B}}_{p+1}(F),\hat{\mathbf{C}}(F))$$

in which the plug-in estimator of the unknown parameter $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}} = \mathbf{v}^{1/(p+1)}(\hat{\mathbf{B}}_{p+1}(F),\hat{\mathbf{C}}(F))$, is computed by the Newton–Raphson method using the gradient and Hessian formulae of Theorem 4 and where the plug-in estimators of the unknown matrices $\mathbf{B}_{p+1}(F) = \{B_{\alpha\beta,p+1}(F)\}_{\alpha,\beta=1}^d$, $\mathbf{C}(F)$ are

$$\begin{aligned}\hat{\mathbf{B}}_{p+1}(F) &= \{\hat{B}_{\alpha\beta,p+1}(F)\}_{\alpha,\beta=1}^d, \quad \hat{\mathbf{C}}(F) = \{\hat{C}_\alpha(F)\}_{\alpha=1}^d,\\ \hat{B}_{\alpha\beta,p+1}(F) &= n^{-1}\sum_{j=1}^n\left\{n^{-1}\sum_{i=1}^n K_{g_\alpha}^{(p)}(X_{j\alpha}-X_{i\alpha})\prod_{\gamma=1,\gamma\neq\alpha}^d\int_{-\infty}^{X_{j\gamma}}K_{g_\gamma}(x_\gamma - X_{i\gamma})dx_\gamma\right\}\\ &\quad\times\left\{n^{-1}\sum_{i=1}^n K_{g_\beta}^{(p)}(X_{j\beta}-X_{i\beta})\prod_{\gamma=1,\gamma\neq\beta}^d\int_{-\infty}^{X_{i\gamma}}K_{g_\gamma}(x_\gamma - X_{i\gamma})dx_\gamma\right\},\\ \hat{C}_\alpha(F) &= n^{-1}\sum_{j=1}^n\left\{n^{-1}\sum_{i=1}^n K_{g_\alpha}(X_{j\alpha}-X_{i\alpha})\prod_{\gamma=1,\gamma\neq\alpha}\int_{-\infty}^{X_{j\gamma}}K_{g_\gamma}(x_\gamma - X_{i\gamma})dx_\gamma\right\}.\end{aligned}$$

The pilot bandwidth vector $\mathbf{g} = (g_1,\ldots,g_d)^{\mathrm{T}}$ is the simple rule-of-thumb bandwidth for multivariate density estimation in Scott [9].

In the next section, we present Monte Carlo evidence for Theorems 2 and 3, and illustrate the use of the smooth estimator $\hat{F}(\mathbf{x})$ with real data examples.

4. Examples

In all computing of this section, we use the quartic kernel $K(u) = 15/16 \times (1-u^2)^2 I(|u| \le 1)$ with $p = 1$ and the plug-in bandwidth vector $\hat{\mathbf{h}}_{\text{opt}}$ described in the previous section. We have not experimented with other choices of K and p due to limit of space and as these choices are in general not as crucial as that of the bandwidth, see Fan and Yao [10].

4.1. *A simulated example*

We examine in this subsection the asymptotic results of Theorems 2 and 3 via simulation. The data are generated from the following vector autoregression (VAR) equation

$$\mathbf{X}_t = a\mathbf{X}_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \Sigma), \quad 2 \le t \le n, \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad 0 \le a, \quad \rho < 1$$

with stationary distribution $\mathbf{X}_t = (X_{t1}, X_{t2})^{\text{T}} \sim N(0, (1-a^2)^{-1}\Sigma)$. Clearly, higher values of a correspond to stronger dependence among the observations, and in particular, if $a = 0$, the data is i.i.d. The parameter ρ controls the orientation of the bivariate cdf F, and in particular, if $a = \rho = 0$, then F is a bivariate standard normal distribution. To cover various scenarios, we have experimented with three cases: $\rho = 0, a = 0$; $\rho = 0.5, a = 0.2$; $\rho = 0.9, a = 0.2$.

A total of 100 samples $\{\mathbf{X}_t\}_{t=1}^n$ of sizes $n = 50, 100, 200, 500$ are generated, and $\hat{F}$ is computed using the optimal bandwidth vector $\hat{\mathbf{h}}_{\text{opt}}$ described in Section 3. Of interest are the means over the 100 replications of the global maximal deviation $D_n(\hat{F})$ defined in Equation (3), denoted as $\bar{D}_n(\hat{F})$, and the MISE$(\hat{F}; \hat{\mathbf{h}}_{\text{opt}})$ defined in Equation (2). Both measures are listed in Table 1. As one examines in Table 1, both $\bar{D}_n(\hat{F})$ and MISE$(\hat{F}; \hat{\mathbf{h}}_{\text{opt}})$ values decrease as sample size increases in all cases, corroborating with Theorems 2 and 3. Also listed in Table 1 are the differences of the same measures for the empirical cdf $\widehat{F}$ against those of $\hat{F}$, which are always positive regardless of the data generating process (*i.e.* for different combinations of a, ρ) and measures of deviation (*i.e.* $\bar{D}_n$ or MISE). This corroborates with Theorem 5 that $\hat{F}$ has asymptotically smaller MISE than $\widehat{F}$.

Table 1. $\bar{D}_n$ and mean integrated squared error (MISE) of $\hat{F}$ and $\widehat{F}$.

	n	$\bar{D}_n(\hat{F})$	$\bar{D}_n(\widehat{F}) - \bar{D}_n(\hat{F})$	MISE$(\hat{F})$	MISE$(\widehat{F})$ − MISE$(\hat{F})$
$\rho = 0, a = 0$	50	0.10137	0.05528	0.15751	0.02100
	100	0.07385	0.0352	0.07202	0.01046
	200	0.05107	0.02292	0.03350	0.00411
	500	0.03482	0.0122	0.01421	0.00184
$\rho = 0.5, a = 0.2$	50	0.10761	0.05109	0.20178	0.03239
	100	0.07514	0.03423	0.08897	0.01539
	200	0.05271	0.02204	0.04169	0.00409
	500	0.03714	0.01121	0.01936	0.00236
$\rho = 0.9, a = 0.2$	50	0.10687	0.03561	0.20294	0.03549
	100	0.07339	0.0243	0.08635	0.01479
	200	0.05038	0.01575	0.04008	0.00628
	500	0.03668	0.00837	0.02028	0.00215

Based on these observations, we believe our kernel estimator of multivariate cdf is a convenient and reliable tool, which is also superior to the empirical cdf in terms of accuracy.

4.2. *GDP growth and unemployment*

In this subsection, we discuss in detail the dependence of the US GDP quarterly growth rate on the unemployment rate. There are three types of unemployment: frictional, structural and cyclical. Economists regard frictional and structural unemployment as essentially unavoidable in a dynamic economy; so full employment is something less than 100% employment. The full employment rate of unemployment is also referred to as the natural rate of unemployment. It does not mean that the economy will always operate at the natural rate. The economy sometimes operates at an unemployment rate higher than the natural rate due to cyclical unemployment. In contrast, the economy may on some occasions achieve an unemployment rate below the natural rate.

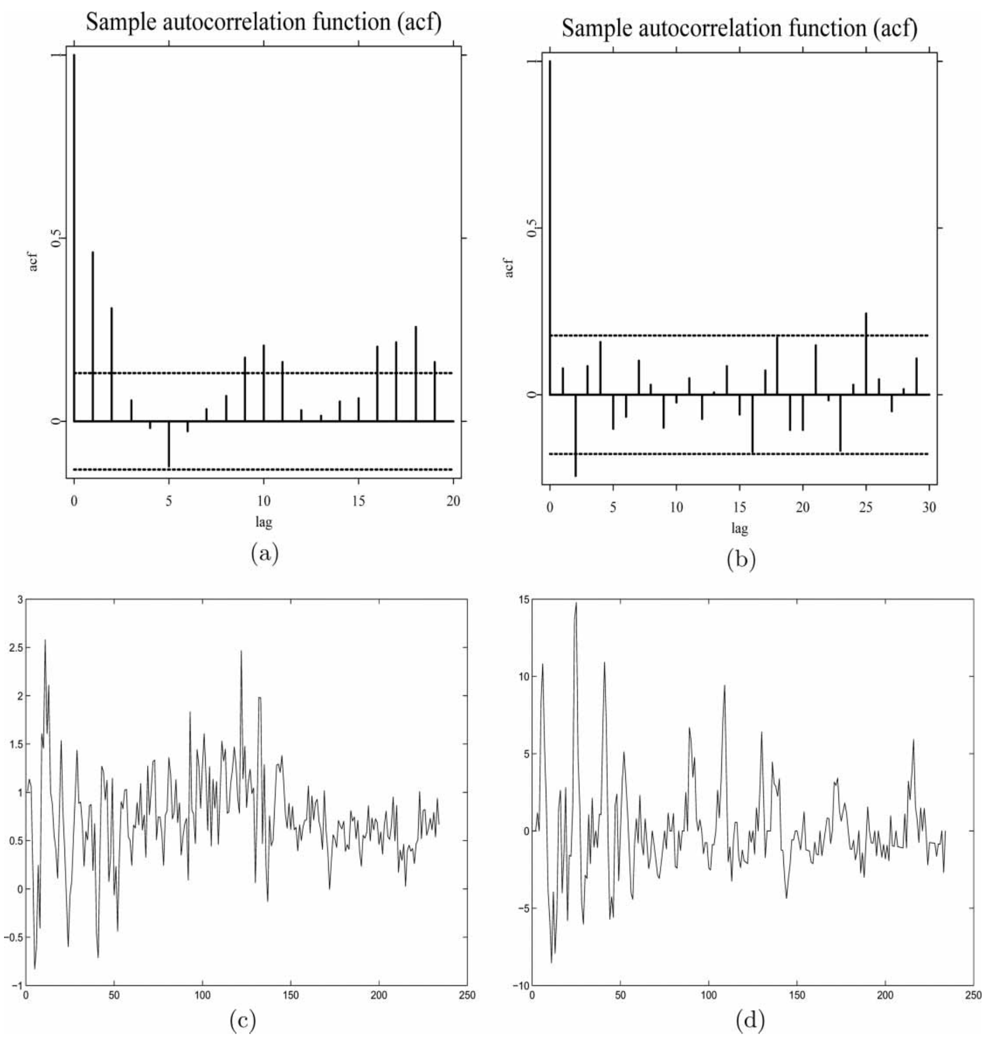

Figure 1. (a) ACF plot of gross domestic product (GDP) quarterly growth rate; (b) ACF plot of unemployment quarterly growth rate; (c) time plot of GDP quarterly growth rate; (d) time plot of unemployment quarterly growth rate. In all ACF plots, the horizontal lines are at 0 and at $\pm 1.96n^{-1/2}$, the 95% confidence limits.

For example, during World War II, when the natural rate was about 4%, actual rate fell below 2% during 1943–1945. The pressure of wartime production resulted in an almost unlimited demand for labour. The natural rate is not forever fixed. It was about 4% in the 1960s, and economists generally agreed that the natural rate was about 6%. Today, the consensus is that the rate is about 5.5%.

GDP gap denotes the amount by which actual GDP falls short of the theoretical GDP under the natural rate. Okun's law, based on recent estimates, indicates that for every 1% by which the actual unemployment rate exceeds the natural rate, a GDP gap of about 2% occurs. See Samuelson [11, p. 559] or McConnell and Brue [12, p. 214] for more details. In other words, if unemployment rate falls, then GDP growth rate increases. But unemployment rate cannot keep falling because it moves around the natural rate. So it is useful to find the relationship between the GDP growth rate and unemployment growth rate.

Let X_{t1} be the seasonally adjusted quarterly unemployment growth rate in quarter t, X_{t2} be the quarterly GDP growth rate in quarter t, all data taken from the first quarter of 1948 ($t = 1$) to the second quarter of 2006 ($t = 234$). Since all data has been seasonally adjusted, it is reasonable to treat $\mathbf{X}_t = (X_{t1}, X_{t2})^{\mathrm{T}}, t = 1, \ldots, 234$ as a strictly stationary time series, which is shown in the time plots. The ACF plots indicate that the autocorrelation function does not deviate significantly from geometric decay, which is a consequence of the geometric α-mixing Assumption (A2). The plots are shown in Figure 1.

Given any interval $I = [a, b]$, the survival function of X_{t2} conditional on $X_{t1} \in I$ is defined as

$$S_I(x_2) = P(X_{t2} > x_2 | X_{t1} \in I) = 1 - \frac{F(b, x_2) - F(a, x_2)}{F(b, +\infty) - F(a, +\infty)} \tag{4}$$

in which F is the joint distribution function of X_{t1} and X_{t2}.

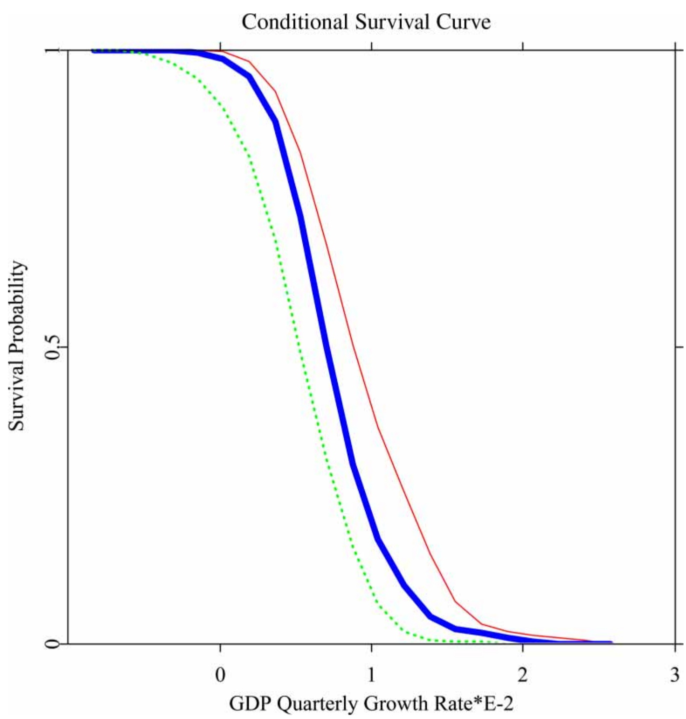

Figure 2. Survival curves of gross domestic product growth rate conditional on unemployment growth rate: $X_{t1} \in [-0.08, -0.04]$, thin solid; $X_{t1} \in [-0.02, 0.02]$, thick solid; $X_{t1} \in [0.04, 0.08]$, dotted.

The function $S_I(x_2)$ can be approximated by the following plug-in estimator

$$\hat{S}_I(x_2) = 1 - \frac{\hat{F}(b, x_2) - \hat{F}(a, x_2)}{\hat{F}(b, +\infty) - \hat{F}(a, +\infty)} \tag{5}$$

in which $\hat{F}$ is the kernel estimator of F defined in Equation (1). According to Theorems 1 and 3, for any fixed x_2, $|\hat{S}_I(x_2) - S_I(x_2)| = O_p(n^{-1/2})$ while

$$\sup_{x_2 \in R} |\hat{S}_I(x_2) - S_I(x_2)| = O_{a.s.}(n^{-1/2} \log n),$$

so the estimator $\hat{S}_I(x_2)$ is theoretically very reliable. We therefore draw probabilistic conclusions based on the smooth estimate $\hat{S}_I(x_2)$ instead of the true $S_I(x_2)$.

In Figure 2, the estimated conditional survival curve $\hat{S}_I(x_2)$ is plotted for intervals $I = [-0.08, -0.04]$, $I = [-0.02, 0.02]$ and $I = [0.04, 0.08]$. Clearly, when the unemployment

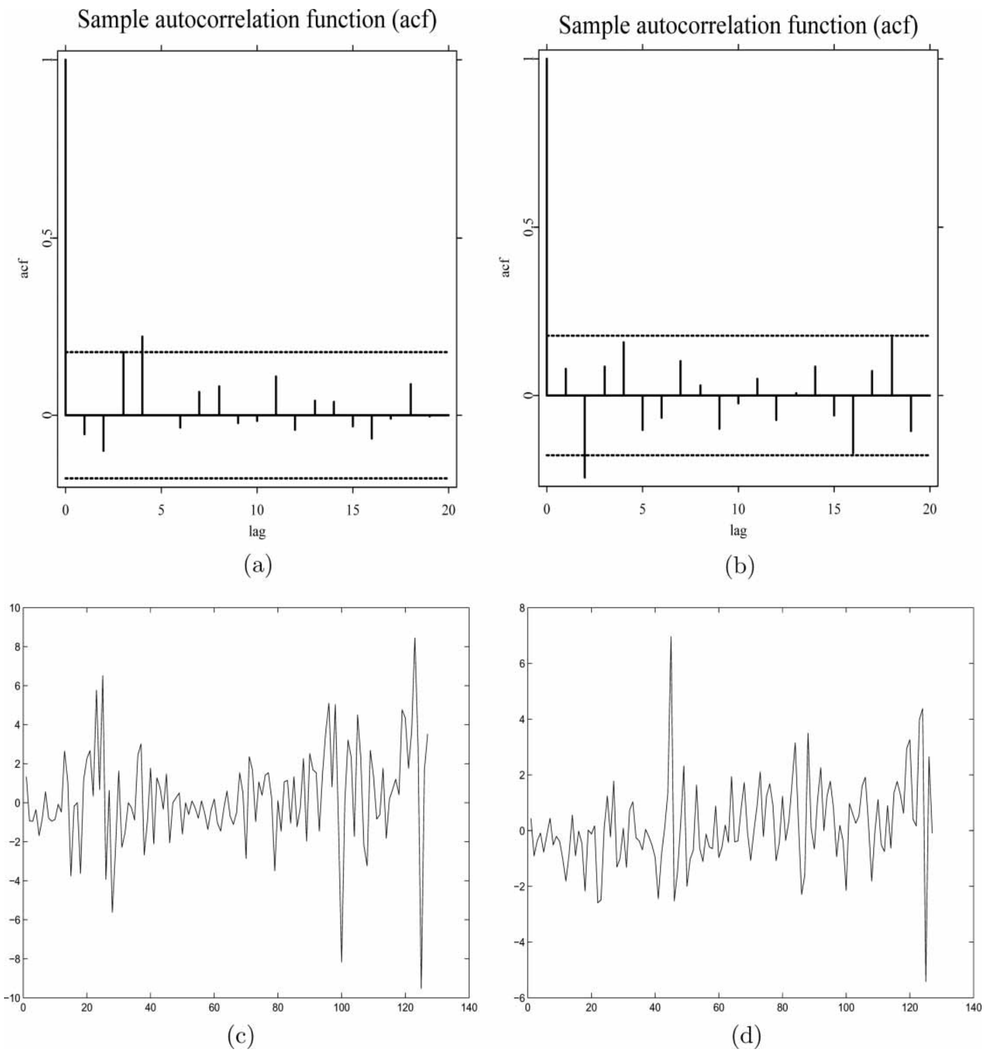

Figure 3. (a) ACF plot of gold price return; (b) ACF plot of silver price return; (c) time plot of gold price return; (d) time plot of silver price return. In all ACF plots, the horizontal lines are at 0 and at $\pm 1.96n^{-1/2}$, the 95% confidence limits.

growth rate is between −0.08 and −0.04, the chance to have the GDP growth rate higher than 1.5% is the greatest, which is about 0.2. This is in accordance with the Okun's law that the growth in GDP is associated with the unemployment rate. So if policy-makers want to achieve a high GDP growth rate, they should find better ways to lower the unemployment rate. One can even estimate the probabilities of GDP growth rates given the policy of unemployment, which is the interval I. If current unemployment rate is close to the natural rate, then the I is an interval close to 0, such as $[-0.02, 0.02]$; if the current unemployment rate is much higher than the natural rate, then the I is a negative interval, *i.e.* trying to lower the unemployment rate.

On the other hand, the survival function of X_{t1} conditional on X_{t2} can be computed similarly. If a certain level of GDP growth rate is planned to be achieved, one can estimate the conditional probabilities of different unemployment growth rates.

4.3. *Gold and silver price returns*

In this subsection, we discuss in detail the dependence of price returns of gold on silver. Let X_{t1} be the monthly silver price return in quarter t, X_{t2} be the monthly gold price return in quarter t, all data taken from the February of 1996 ($t = 1$) to the August of 2006 ($t = 127$). Since both data have been seasonally adjusted, it is reasonable to treat $\mathbf{X}_t = (X_{t1}, X_{t2})^{\mathrm{T}}, t = 1, \ldots, 127$ as a strictly stationary time series, which is shown in the time plots. Again, the ACF plots indicate that the autocorrelation function does not deviate significantly from geometric decay, which is a consequence of the geometric α-mixing Assumption (A2). The plots are shown in Figure 3.

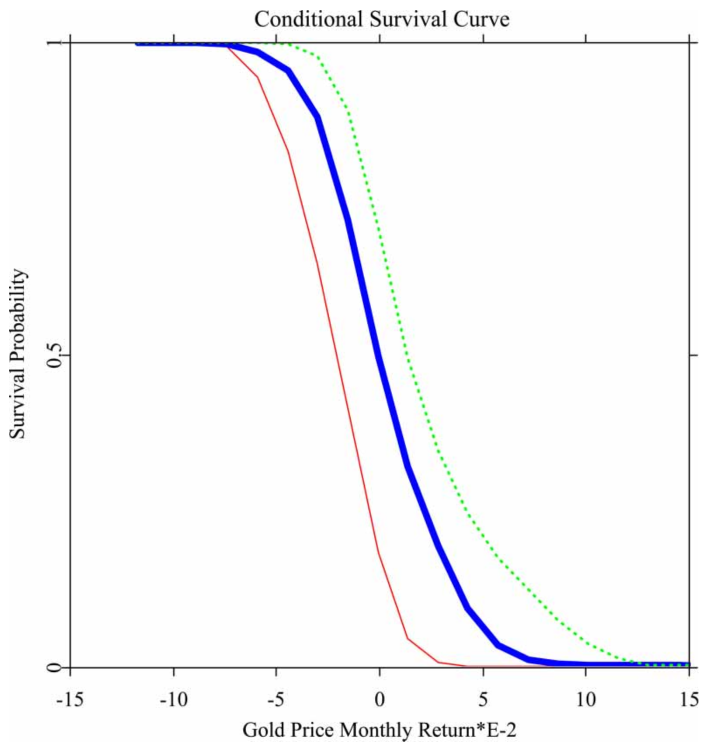

Figure 4. Survival curves of gold price return conditional on silver price return: $X_{t1} \in [-0.10, -0.06]$, thin solid; $X_{t1} \in [-0.02, 0.02]$, thick solid; $X_{t1} \in [0.06, 0.10]$, dotted.

Similar to the previous example, for any fixed interval $I = [a, b]$, the survival function $S_I(x_2)$ and its estimate $\hat{S}_I(x_2)$ are defined as in Equations (4) and (5), respectively. We again base our inference on the estimated function $\hat{S}_I(x_2)$.

In Figure 4, the estimated conditional survival curve $\hat{S}_I(x_2)$ is plotted for intervals $I = [-0.10, -0.06]$, $I = [-0.02, 0.02]$, $I = [0.06, 0.10]$. Clearly, when the silver price return is higher, the gold price increases faster. This is in accordance with the economic theory of substitute goods, *i.e.* increase in the price of one good causes increases of demand of other substitutes, hence the increases of the prices of substitutes. So gold and silver clearly substitute each other. See Samuelson [11, p. 81] for more details.

On the other hand, the survival function of X_{t1} conditional on X_{t2} can be computed similarly. That is the conditional probability of silver price return based on gold price return.

Acknowledgements

This research is part of the first author's dissertation under the supervision of the second author, and has been supported in part by NSF awards, DMS 0405330 and DMS 0706518. The authors are grateful to two referees and editor Michael Akritas for their constructive comments that led to the significant improvement of this paper.

References

[1] Q. Li and J.S. Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton, 2007.
[2] H. Yamato, *Uniform convergence of an estimator of a distribution function*, Bull. Math. Statist. 15 (1973), pp. 69–78.
[3] P.J. Bickel and M. Rosenblatt, *On some global measures of the deviations of density function estimates*, Ann. Statist. 1 (1973), pp. 1071–1095.
[4] R.D. Reiss, *Nonparametric estimation of smooth distribution functions*, Scand. J. Statist. 8 (1981), pp. 116–119.
[5] M. Falk, *Asymptotic normality of kernel type estimators of quantiles*, Ann. Statist. 13 (1985), pp. 428–433.
[6] M.Y. Cheng and L. Peng, *Regression modeling for nonparametric estimation of distribution and quantile functions*, Statist. Sinica 12 (2002), pp. 1043–1060.
[7] D. Bosq, *Nonparametric Statistics for Stochastic Processes*, Springer-Verlag, New York, 1998.
[8] L. Yang and R. Tschernig, *Multivariate bandwidth selection for local linear regression*, J. R. Statist. Soc., Ser. B 61 (1999), pp. 793–815.
[9] D.W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons, New York, 1992.
[10] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer-Verlag, New York, 2003.
[11] P. Samuelson, *Economics*, Irwin/McGraw-Hill, New York, 1995.
[12] C. McConnell and S. Brue, *Economics: Principles, Problems, and Policies*, Irwin/McGraw-Hill, Boston, 1999.
[13] J. Sunklodas, *On the rate of convergence in the central limit theorem for strongly mixing random variables*, Lithuanian Math. J. 24 (1984), pp. 182–190.

Appendix

A1. Preliminaries

In this appendix, we denote by C (or c) any positive constants, by U (or u) sequences of random variables that are uniformly O (or o) of certain order and by $O_{a.s.}$ almost surely O, etc.

LEMMA A1 (Berry–Esseen inequality, [13, Theorem 1]) *Let $\{\xi_i\}_{i=1}^n$ be an α-mixing sequence with $E\xi_n = 0$. Denote $d_\delta := \max_{1\le i\le n}\{E|\xi_i|^{2+\delta}\}$, $0 < \delta \le 1$, $S_n = \sum_{i=1}^n \xi_i$, $\sigma_n^2 := ES_n^2 \ge c_0 n$ for some $c_0 \in (0, +\infty)$. If $\alpha(n) \le K_0 \exp(-\lambda_0 n)$, $\lambda_0 > 0$, $K_0 > 0$, then there exist $c_1 = c_1(K_0, \delta)$, $c_2 = c_2(K_0, \delta)$, such that*

$$\Delta_n = \sup_z |P\{\sigma_n^{-1} S_n < z\} - \Phi(z)| \le c_1 \frac{d_\delta}{c_0 \sigma_n^\delta} \left\{ \log\left(\frac{\sigma_n}{c_0^{1/2}}\right)/\lambda \right\}^{1+\delta} \tag{A1}$$

for any λ with $\lambda_1 \le \lambda \le \lambda_2$, where

$$\lambda_1 = c_2 \frac{\{\log(\sigma_n/c_0^{1/2})\}^b}{n}, \quad \frac{b > 2(1+\delta)}{\delta} \quad \lambda_2 = 4(2+\delta)\delta^{-1} \log\left(\frac{\sigma_n}{c_0^{1/2}}\right).$$

LEMMA A2 (Bernstein's inequality, [7, Theorem 1.4]) *Let* $\{\xi_t\}$ *be a zero mean real-valued process,* $S_n = \sum_{i=1}^n \xi_i$. *Suppose that there exists* $c > 0$ *such that for* $i = 1, \ldots, n$, $k \geq 3, E|\xi_i|^k \leq c^{k-2}k!E\xi_i^2 < +\infty$, $m_r = \max_{1\leq i\leq N} \|\xi_i\|_r$, $r \geq 2$. *Then for each* $n > 1$, *integer* $q \in [1, n/2]$, *each* $\varepsilon > 0$ *and* $k \geq 3$

$$P\left\{\left|\sum_{i=1}^n \xi_i\right| > n\varepsilon_n\right\} \leq a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) + a_2(k)\alpha\left(\left[\frac{n}{q+1}\right]\right)^{2k/(2k+1)}$$

where

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right), \quad a_2(k) = 11n\left(1 + \frac{5m_k^{2k/(2k+1)}}{\varepsilon_n}\right).$$

A2. Proofs of Theorems 1 and 2

LEMMA A3 *Under Assumptions* (A1), (A3) and (A4), *as* $n \to \infty$

$$E\{\hat{F}(\mathbf{x})\} = F(\mathbf{x}) + \frac{\mu_{p+1}(K)}{(p+1)!}\sum_{\alpha=1}^d h_\alpha^{p+1}\frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}} + u(h_{\max}^{p+1}).$$

Proof Using the integral form of Taylor expansion and denoting $\mathbf{hv} = (h_1v_1, \ldots, h_dv_d)^{\mathrm{T}}$, we write

$$f(\mathbf{u} + \mathbf{hv}) \equiv f(\mathbf{u}) + \sum_{r=1}^p \frac{1}{r!}\left(\sum_{\alpha=1}^d h_\alpha v_\alpha \frac{\partial}{\partial u_\alpha}\right)^r f(\mathbf{u}) + R_{p+1},$$

$$R_{p+1} = R_{p+1}(\mathbf{u}, \mathbf{hv}) = \int_0^1 \left\{\frac{t^p}{p!}\left(\sum_{\alpha=1}^d h_\alpha v_\alpha \frac{\partial}{\partial u_\alpha}\right)^{p+1} f(\mathbf{u} + t\mathbf{hv})\right\} dt.$$

Hence Assumptions (A4), (A1) and (A3) sequentially imply that

$$\begin{aligned}
E\{\hat{F}(\mathbf{x})\} &= E\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u} = \int_{-\infty}^{\mathbf{x}} d\mathbf{u}\int_{[-1,1]^d} f(\mathbf{u} + \mathbf{hv})K(\mathbf{v})d\mathbf{v} \\
&= \int_{-\infty}^{\mathbf{x}} f(\mathbf{u})d\mathbf{u} + \int_{-\infty}^{\mathbf{x}} d\mathbf{u}\int_{[-1,1]^d}\left[\sum_{r=1}^p \frac{1}{r!}\left(\sum_{\alpha=1}^d h_\alpha v_\alpha \frac{\partial}{\partial u_\alpha}\right)^r f(\mathbf{u}) + R_{p+1}\right] K(\mathbf{v})d\mathbf{v} \\
&= F(9\mathbf{x}) + \int_{-\infty}^{\mathbf{x}} d\mathbf{u}\int_{[-1,1]^d}\left[\int_0^1\left\{\frac{t^p}{p!}\left(\sum_{\alpha=1}^d h_\alpha v_\alpha \frac{\partial}{\partial u_\alpha}\right)^{p+1} f(\mathbf{u} + t\mathbf{hv})\right\} dt\right] K(\mathbf{v})d\mathbf{v} \\
&= F(\mathbf{x}) + \frac{\mu_{p+1}(K)}{(p+1)!}\int_{-\infty}^{\mathbf{x}}\sum_{\alpha=1}^d h_\alpha^{p+1}\frac{\partial^{p+1}f}{\partial u_\alpha^{p+1}}(\mathbf{u})d\mathbf{u} + u(h_{\max}^{p+1}) \\
&= F(\mathbf{x}) + \frac{\mu_{p+1}(K)}{(p+1)!}\sum_{\alpha=1}^d h_\alpha^{p+1}\frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}} + u(h_{\max}^{p+1}).
\end{aligned}$$

■

LEMMA A4 *Under Assumptions* (A1)–(A4), *as* $n \to \infty$

$$E\left\{\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_j - \mathbf{u})d\mathbf{u}\right\} = \begin{cases} F(\mathbf{x}) - D(K)\sum_{\alpha=1}^d h_\alpha \dfrac{\partial F(\mathbf{x})}{\partial x_\alpha} + u(h_{\max}) & i = j, \\ EI\{\mathbf{X}_i \leq \mathbf{x}\}I\{\mathbf{X}_j \leq \mathbf{x}\} + u(h_{\max}) & i \neq j. \end{cases}$$

Proof We begin with the case of $i = j$,

$$E\left\{\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}\right\}^2 = \int_{-\infty}^{\infty} f(\mathbf{v})\tilde{K}\left(\frac{\mathbf{x}-\mathbf{v}}{\mathbf{h}}\right)^2 d\mathbf{v} = \int_{-1}^{\infty} f(\mathbf{x}-\mathbf{hw})\tilde{K}^2(\mathbf{w})h_{\text{prod}}d\mathbf{w}$$

$$= h_{\text{prod}}\int_{-1}^{\infty}\{I(\mathbf{w} \geq -1) - I(\mathbf{w} \geq 1)\}f(\mathbf{x}-\mathbf{hw})\tilde{K}^2(\mathbf{w})d\mathbf{w} + \int_{1}^{\infty} f(\mathbf{x}-\mathbf{hw})h_{\text{prod}}d\mathbf{w}$$

$$= h_{\text{prod}}\int_{-1}^{\infty}\{I(\mathbf{w} \geq -1) - I(\mathbf{w} \geq 1)\}f(\mathbf{x}-\mathbf{hw})\tilde{K}^2(\mathbf{w})d\mathbf{w} + F(\mathbf{x}-\mathbf{h})$$

$$= \sum_{\alpha=1}^{d} h_{\text{prod}}\int_{\mathbf{1}}^{\infty} d\mathbf{w}_{\_\alpha}\int_{-1}^{1} dw_\alpha f(\mathbf{x}-\mathbf{hw})\tilde{K}^2(w_\alpha) + F(\mathbf{x}) - \sum_{\alpha=1}^{d}\frac{\partial F(\mathbf{x})}{\partial x_\alpha}h_\alpha + u(h_{\max})$$

$$= F(\mathbf{x}) - \sum_{\alpha=1}^{d} h_\alpha \frac{\partial F(\mathbf{x})}{\partial x_\alpha}D(K) + u(h_{\max}).$$

Similarly, for the case of $i \neq j$, one obtains

$$E\left\{\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_j - \mathbf{u})d\mathbf{u}\right\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{v}_i d\mathbf{v}_j f_{i,j}(\mathbf{v}_i, \mathbf{v}_j)\tilde{K}\left(\frac{\mathbf{x}-\mathbf{v}_i}{\mathbf{h}}\right)\tilde{K}\left(\frac{\mathbf{x}-\mathbf{v}_j}{\mathbf{h}}\right)$$

$$= \int_{-1}^{\infty}\int_{-1}^{\infty} f_{i,j}(\mathbf{x}-\mathbf{hw}_i, \mathbf{x}-\mathbf{hw}_j)\tilde{K}(\mathbf{w}_i)\tilde{K}(\mathbf{w}_J)h_{\text{prod}}^2 d\mathbf{w}_i d\mathbf{w}_j$$

$$= h_{\text{prod}}^2\left\{\int_{-1}^{\infty}\{I(\mathbf{w}_i \geq -1) - I(\mathbf{w}_i \geq 1)\}\tilde{K}(\mathbf{w}_i)d\mathbf{w}_i + \int_{1}^{\infty} d\mathbf{w}_i\right\}$$

$$\times\left\{\int_{-1}^{\infty}\left\{I(\mathbf{w}_j \geq -1) - I(\mathbf{w}_j \geq 1)\right\}\tilde{K}(\mathbf{w}_j)d\mathbf{w}_j + \int_{1}^{\infty} d\mathbf{w}_j\right\} f_{i,j}(\mathbf{x}-\mathbf{hw}_i, \mathbf{x}-\mathbf{hw}_j)$$

$$= h_{\text{prod}}^2\int_{-1}^{\infty}\{I(\mathbf{w}_i \geq -1) - I(\mathbf{w}_i \geq 1)\}\tilde{K}(\mathbf{w}_i)d\mathbf{w}_i\int_{1}^{\infty} d\mathbf{w}_j f_{i,j}(\mathbf{x}-\mathbf{hw}_i, \mathbf{x}-\mathbf{hw}_j)$$

$$+ h_{\text{prod}}^2\int_{-1}^{\infty}\{I(\mathbf{w}_j \geq -1) - I(\mathbf{w}_j \geq 1)\}\tilde{K}(\mathbf{w}_j)d\mathbf{w}_j\int_{1}^{\infty} d\mathbf{w}_i f_{i,j}(\mathbf{x}-\mathbf{hw}_i, \mathbf{x}-\mathbf{hw}_j)$$

$$+ EI\{\mathbf{X}_i \leq \mathbf{x}-\mathbf{h}\}I\{\mathbf{X}_j \leq \mathbf{x}-\mathbf{h}\} + u(h_{\max})$$

$$= \sum_{\alpha=1}^{d} h_\alpha\int_{\text{h}}^{\infty} d\mathbf{v}_j\int_{\mathbf{h}_{i\_\alpha}}^{\infty} d\mathbf{v}_{i\_\alpha}\int_{-1}^{1}\tilde{K}(w_{i\alpha})dw_{i\alpha}f_{i,j}(x_\alpha - hw_{i\alpha}, \mathbf{x}_{\_\alpha} - \mathbf{v}_{i\_\alpha}, \mathbf{x}-\mathbf{v}_j)$$

$$+ \sum_{\alpha=1}^{d} h_\alpha\int_{\text{h}}^{\infty} d\mathbf{v}_i\int_{\mathbf{h}_{j\_\alpha}}^{\infty} d\mathbf{v}_{j\_\alpha}\int_{-1}^{1}\tilde{K}(w_{j\alpha})dw_{j\alpha}f_{i,j}(\mathbf{x}-\mathbf{v}_i, x_\alpha - hw_{j\alpha}, \mathbf{x}_{\_\alpha} - \mathbf{v}_{j\_\alpha})$$

$$+ EI\{\mathbf{X}_i \leq \mathbf{x}\}I\{\mathbf{X}_j \leq \mathbf{x}\} - \sum_{\alpha=1}^{d} h_\alpha\frac{\partial EI\{\mathbf{X}_i \leq \mathbf{x}\}I\{\mathbf{X}_j \leq \mathbf{x}\}}{\partial x_\alpha} + u(h_{\max})$$

$$= EI\{\mathbf{X}_i \leq \mathbf{x}\}I\{\mathbf{X}_j \leq \mathbf{x}\} - \sum_{\alpha=1}^{d} h_\alpha\frac{\partial EI\{\mathbf{X}_i \leq \mathbf{x}\}I\{\mathbf{X}_j \leq \mathbf{x}\}}{\partial x_\alpha}$$

$$+ \sum_{\alpha=1}^{d} h_\alpha\frac{\partial EI\{\mathbf{X}_i \leq \mathbf{x}\}I\{\mathbf{X}_j \leq \mathbf{x}\}}{\partial x_\alpha} + u(h_{\max}) = EI\{\mathbf{X}_i \leq \mathbf{x}\}I\{\mathbf{X}_j \leq \mathbf{x}\} + u(h_{\max}). \quad \blacksquare$$

Denote $S_n = S_n(\mathbf{x}) = n\{\hat{F}(\mathbf{x}) - E\hat{F}(\mathbf{x})\} = \sum_{i=1}^{n}\xi_{in}$ in which

$$\xi_{i,n} = \xi_{i,n}(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u} - E\left\{\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}\right\},$$

then clearly $E\xi_{i,n} = 0$. Denote by $\tilde{\gamma}(l) = \text{cov}(\xi_{i,n}, \xi_{i+l,n})$ the autocovariance function, then we have the following corollary.

COROLLARY A1 *Under Assumptions* (A1)–(A4), *as* $n \to \infty$

$$\text{cov}(\xi_{i,n}, \xi_{j,n}) = \tilde{\gamma}(i-j) = \begin{cases} F(\mathbf{x}) - F^2(\mathbf{x}) - D(K)\sum_{\alpha=1}^{d} h_\alpha \dfrac{\partial F(\mathbf{x})}{\partial x_\alpha} + u(h_{\max}) & i = j, \\ EI\{\mathbf{X}_i \le \mathbf{x}\}I\{\mathbf{X}_j \le \mathbf{x}\} - F^2(\mathbf{x}) + u(h_{\max}) & i \ne j. \end{cases}$$

Proof According to Lemmas A3 and A4,

$$\begin{aligned} \text{cov}(\xi_{i,n}, \xi_{j,n}) &= \left[E\left\{\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u} \int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_j - \mathbf{u})d\mathbf{u}\right\} - \left(E\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}\right)^2\right] \\ &= \begin{cases} F(\mathbf{x}) - D(K)\sum_{\alpha=1}^{d} h_\alpha \frac{\partial F(\mathbf{x})}{\partial x_\alpha} + u(h_{\max}) & i = j \\ EI\{\mathbf{X}_i \le \mathbf{x}\}I\{\mathbf{X}_j \le \mathbf{x}\} + u(h_{\max}) & i \ne j \end{cases} \\ &\quad - \left[F(\mathbf{x}) + \frac{\mu_{p+1}(K)}{(p+1)!}\sum_{\alpha=1}^{d} h_\alpha^{p+1} \frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}} + u(h_{\max}^{p+1})\right]^2, \end{aligned}$$

the rest of the proof is trivial. ■

Proof of Theorems 1 and 2 According to Corollary A1

$$\tilde{\gamma}(l) = \begin{cases} \gamma(0) - D(K)\sum_{\alpha=1}^{d} h_\alpha \dfrac{\partial F(\mathbf{x})}{\partial x_\alpha} + u(h_{\max}) & l = 0, \\ \gamma(l) + u(h_{\max}) & l \ne 0, \end{cases} \tag{A2}$$

in which $\gamma(l) = \gamma(l, \mathbf{x}) = EI\{\mathbf{X}_1 \le \mathbf{x}\}I\{\mathbf{X}_{1+l} \le \mathbf{x}\} - F^2(\mathbf{x})$. Lemma A3 and Assumption (A3) further imply that

$$S_n = n\left\{\hat{F}(\mathbf{x}) - F(\mathbf{x}) - \frac{\mu_{p+1}(K)}{(p+1)!}\sum_{\alpha=1}^{d} h_\alpha^{p+1} \frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}} + u(h_{\max}^{p+1})\right\}. \tag{A3}$$

Meanwhile, $\sigma_n^2 = ES_n^2 = \text{var}(S_n) = nA_n + nB_n$, where $A_n = \sum_{|l| \le c\log n}(1 - |l|/n)\tilde{\gamma}(l)$ and $B_n = \sum_{c\log n < |l| < n}(1 - |l|/n)\tilde{\gamma}(l)$. Because $|\gamma(l)|$ is

$$|P(\{\omega : \mathbf{X}_1(\omega) \le \mathbf{x}\} \cap \{\omega : \mathbf{X}_{1+h}(\omega) \le \mathbf{x}\}) - P(\{\omega : \mathbf{X}_1(\omega) \le \mathbf{x}\})P(\{\omega : \mathbf{X}_{1+h}(\omega) \le \mathbf{x}\})|$$

which is bounded by $\alpha(l) \le K_0 e^{-\lambda_0 l}$. Then, $\sum_{l=-\infty}^{\infty} |\gamma(l)| \le \gamma(0) + 2\sum_{l=1}^{\infty} K_0 \exp(-\lambda_0 l) < \infty$ and Equation (A2) implies that

$$A_n = \sum_{|l| \le c\log n}\left(\frac{1-|l|}{n}\right)\gamma(l) + \sum_{|l| \le c\log n}\left(\frac{1-|l|}{n}\right)U(h_{\max}) \to \sum_{l=-\infty}^{\infty}\gamma(l) \ge c_0.$$

Next, $|\text{cov}(\xi_{1,n}, \xi_{(1+l),n})| \le 4\|\xi_{1,n}\|_\infty \|\xi_{(1+l),n}\|_\infty \, \alpha(h) \le 4K_0 \exp(-\lambda_0 l)$ gives

$$|B_n| = \sum_{c\log n < |l| < n}\left(\frac{1-|l|}{n}\right)|\tilde{\gamma}(l)| \le \sum_{|l| > c\log n}\left(\frac{1-|l|}{n}\right)4K_0K_0\exp(-\lambda_0 l).$$

For $c \ge 2/\lambda_0$, $|B_n| \le 4K_0 e^{-\lambda_0 c\log n}/1 - e^{-\lambda_0} = K_0 n^{-c\lambda_0}/1 - e^{-\lambda_0} \le C_1 n^{-2}$. For n large enough, $\sigma_n^2/n = A_n + B_n \to \sum_{l=-\infty}^{\infty}\gamma(l) \ge c_0$, therefore $\sum_{|l| \le n}\gamma(l) > 0$. Then by Equation (A1) in Lemma A1,

$$\Delta_n = \sup_z |P\{\sigma_n^{-1}S_n < z\} - \Phi(z)| \le c_1 \frac{d}{c_0\sigma_n^\delta}\left\{\frac{\log(\sigma_n/c_0^{1/2})}{\lambda}\right\}^{1+\delta}.$$

Let $\delta = 1$, $\lambda = 4(2+\delta)\delta^{-1}\log\left(\sigma_n/c_0^{1/2}\right) = 12\log(\sigma_n/c_0^{1/2})$, $d = 1$, then $\Delta_n \le (c_1)/(c_0\sigma_n)12^{-2} = c/\sigma_n = O(n^{-1/2})$, *i.e.* $S_n/\sigma_n \to_d N(0,1)$. Theorem 1 then follows because $\sqrt{n}\sqrt{V^{-1}(\mathbf{x})}(\hat{F}(\mathbf{x}) - F(\mathbf{x})) \to_d N(0,1)$ by Slutsky's theorem. Equations (A2) and (A3) together with $E\xi_{i,n} = 0$ imply that

$$\{E\hat{F}(\mathbf{x}) - F(\mathbf{x})\}^2 = \frac{\mu_{p+1}^2(K)}{(p+1)!^2}\left\{\sum_{\alpha=1}^{d} h_\alpha^{p+1}\frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}}\right\}^2 + u(h_{\max}^{2p+2}),$$

$$E\{\hat{F}(\mathbf{x}) - E\hat{F}(\mathbf{x})\}^2 = n^{-1}V(\mathbf{x}) - D(K)n^{-1}\sum_{\alpha=1}^{d} h_\alpha \frac{\partial F(\mathbf{x})}{\partial x_\alpha} + u(n^{-1}h_{\max}),$$

hence Theorem 2 follows by computing $\int E\{\hat{F}(\mathbf{x}) - E\hat{F}(\mathbf{x})\}^2 + \{E\hat{F}(\mathbf{x}) - F(\mathbf{x})\}^2 dF(\mathbf{x})$. ■

A3. Proof of Theorem 3

LEMMA A5 *Denote* $g_{m_1,\ldots,m_d} = (a_{1,m_1},\ldots,a_{d,m_d}) \in R^d$, $1 \le m_\alpha \le M_\alpha$ *and*

$$A_n = \max_{1\le m_\alpha \le M_\alpha} |\hat{F}(g_{m_1,\ldots,m_d}) - E\{\hat{F}(g_{m_1,\ldots,m_d})\}|,$$

$$B_n = \max_{1\le m_\alpha \le M_\alpha} |\hat{F}(g_{m_1,\ldots,m_d}) - F(g_{m_1,\ldots,m_d})|.$$

If $\max(M_1,\ldots,M_d) \le Cn$, *then* $A_n + B_n = O_{a.s}(n^{-1/2}\log n)$ *while for i.i.d.* $\mathbf{X}_1,\ldots,\mathbf{X}_n$, $A_n + B_n = O_{a.s.}(n^{-1/2}(\log n)^{1/2})$.

Proof Note that $\hat{F}(g_{m_1,\ldots,m_d}) - E\hat{F}(g_{m_1,\ldots,m_d}) = n^{-1}\sum_{i=1}^n \zeta_{in}$ in which

$$\zeta_{in} = \zeta_{in,m_1,\ldots,m_d} = \zeta_{i,n}(g_{m_1,\ldots,m_d}) = \int_{-\infty}^{g_{m_1,\ldots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u} - E\left\{\int_{-\infty}^{g_{m_1,\ldots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}\right\},$$

then one has $E\zeta_{in} = 0$, while

$$E(\zeta_{in}^2) = E\left(\int_{-\infty}^{g_{m_1,\ldots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u} - E\left\{\int_{-\infty}^{g_{m_1,\ldots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}\right\}\right)^2 \le 1,$$

and for $k \ge 2$, $E(|\zeta_{in}|^k) = E(|\zeta_{in}|^{k-2}\zeta_{in}^2)$, which is

$$E\left[\left|\int_{-\infty}^{g_{m_1,\ldots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u} - E\left\{\int_{-\infty}^{g_{m_1,\ldots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})d\mathbf{u}\right\}\right|^{k-2}\zeta_{in}^2\right] \le 1^{k-2}E(\zeta_{in}^2).$$

By Lemma A2 with $k = 3$, $a_2(3) = 11n(1 + 5m_3^{6/7}/\varepsilon_n)$, $m_2^2 = E(\zeta_{in}^2) \le 1$, $\varepsilon_n = a\log n/\sqrt{n}$,

$$P\left\{\left|\sum_{i=1}^n \zeta_{in}\right| > n\varepsilon_n\right\} \le a_1\exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) + a_2(3)\alpha\left(\left[\frac{n}{(q+1)}\right]\right)^{6/7}.$$

Take q such that $[n/(q+1)] \ge c_0\log n$, $q \ge (c_1 n)/(\log n)$, then $q\varepsilon_n^2/(25m_2^2 + 5c\varepsilon_n) \ge c_2a^2\log n$ and

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) = O(\log n).$$

Since $m_3 = \max_{1\le i\le n}\|\zeta_i\|_3 \le \{E(\zeta_{in}^3)\}^{1/3} \le 1$, then

$$a_2(3) = 11n\left(1 + \frac{5}{\varepsilon_n}\right) \le 11n\left\{1 + \frac{5}{an^{-1/2}\log n}\right\} \le 11n\left\{1 + \frac{5}{a\log n}\right\} = O(n),$$

$$\alpha\left(\left[\frac{n}{(q+1)}\right]\right)^{6/7} \le \left(K_0\exp\left(-\lambda_0\left[\frac{n}{(q+1)}\right]\right)\right)^{6/7} \le Cn^{-6\lambda_0 c_0/7}.$$

So for c_0, c_2 large enough

$$P\left\{\left|\sum_{i=1}^n \zeta_{in}\right| > n\varepsilon_n\right\} \le O(\log n)\exp(-c_2a^2\log n) + Cn^{1-6\lambda_0c_0/7} \le Cn^{-(d+2)},$$

$$P\left\{\max_{1\le m_\alpha\le M_\alpha} n^{-1}\left|\sum_{i=1}^n \zeta_{in,m_1,\ldots,m_d}\right| > an^{-1/2}\log n\right\}$$

$$\le \sum_{m_1=1,\ldots,m_d=1}^{M_1,\ldots,M_d} P\left\{n^{-1}\left|\sum_{i=1}^n \zeta_{in,m_1,\ldots,m_d}\right| > an^{-1/2}\log n\right\} \le Cn^{-(d+2)}\prod_{\alpha=1}^d M_\alpha \le Cn^{-2}.$$

Hence Borel–Cantelli lemma implies that $A_n = O_{a.s.}(n^{-1/2}\log n)$. Meanwhile B_n is bounded by

$$\max_{1\le m_\alpha\le M_\alpha}\left|\hat{F}(g_{m_1,\ldots,m_d}) - E\{\hat{F}(g_{m_1,\ldots,m_d})\}\right| + \max_{1\le m_\alpha\le M_\alpha}\left|E\{\hat{F}(g_{m_1,\ldots,m_d})\} - F(g_{m_1,\ldots,m_d})\right|$$

$$= A_n + U(n^{-1/2}) = O_{a.s.}(n^{-1/2}\log n).$$

If $\mathbf{X}_1,\ldots,\mathbf{X}_n$ are i.i.d., then $A_n + B_n = O_{a.s.}(n^{-1/2}(\log n)^{1/2})$ by using the same steps as shown with Bernstein's inequality of i.i.d. case. ■

LEMMA A6 $\forall A \subset R^d$, $\int_A |K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|d\mathbf{u} \le \int_{R^d} |K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|d\mathbf{u} \le \|K\|_{L^1}^d$.

Proof Applying elementary arguments, $\int_A |K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|d\mathbf{u} \le \int_{R^d} |K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|d\mathbf{u}$ is bounded by

$$\int_{R^d} \left| \prod_{\alpha=1}^{d} h_\alpha^{-1} K\left(\frac{v_\alpha - u_\alpha}{h_\alpha}\right)\right| d\mathbf{u} = \prod_{\alpha=1}^{d} \int_{-1}^{1} |K(w_\alpha)|dw_\alpha \le \|K\|_{L^1}^d. \qquad \blacksquare$$

LEMMA A7 *Let* $-\infty = a_{\alpha,1} < \cdots < a_{\alpha,N_\alpha} = \infty$ *be such that* $\max(N_1,\ldots,N_d) \le Cn$ *and* $P(a_{\alpha,k} \le X_\alpha \le a_{\alpha,k+1}) \le 1/n, \forall 1 \le k \le N_\alpha, \forall 1 \le \alpha \le d$. *Then* $E\int_{g_{n_1,\ldots,n_d}}^{\mathbf{x}} |K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|d\mathbf{u} = u(n^{-1/2}(\log n)^{1/2})$ *in which* $g_{n_1,\ldots,n_d} = (a_{1,n_1},\ldots,a_{d,n_d}) \in R^d$.

Proof

$$\begin{aligned} E\int_{g_{n_1,\ldots,n_d}}^{\mathbf{x}} |K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|d\mathbf{u} &\le \int_{-\infty}^{\infty}\int_{g_{n_1,\ldots,n_d}}^{g_{n_1+1,\ldots,n_d+1}} |K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|d\mathbf{u}\, dF(\mathbf{v}) \\ &= \int_{g_{n_1,\ldots,n_d}-(h_1,\ldots,h_d)}^{g_{n_1+1,\ldots,n_d+1}+(h_1,\ldots,h_d)} dF(\mathbf{v}) \int_{g_{n_1,\ldots,n_d}}^{g_{n_1+1,\ldots,n_d+1}} |K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|d\mathbf{u} \\ &\le C\int_{g_{n_1,\ldots,n_d}-(h_1,\ldots,h_d)}^{g_{n_1+1,\ldots,n_d+1}+(h_1,\ldots,h_d)} dF(\mathbf{v}) \end{aligned}$$

according to Lemma A6. $\int_{g_{n_1,\ldots,n_d}-(h_1,\ldots,h_d)}^{g_{n_1+1,\ldots,n_d+1}+(h_1,\ldots,h_d)} dF(\mathbf{v})$ equals

$$\begin{aligned} &\int_{g_{n_1,\ldots,n_d}-(h_1,\ldots,h_d)}^{g_{n_1+1,\ldots,n_d+1}+(h_1,\ldots,h_d)} dF(\mathbf{v}) - \int_{g_{n_1,\ldots,n_d}}^{g_{n_1+1,\ldots,n_d+1}} dF(\mathbf{v}) + \int_{g_{n_1,\ldots,n_d}}^{g_{n_1+1,\ldots,n_d+1}} dF(\mathbf{v}) \\ &= \int_{g_{n_1,\ldots,n_d}-(h_1,\ldots,h_d)}^{g_{n_1+1,\ldots,n_d+1}+(h_1,\ldots,h_d)} dF(\mathbf{v}) - \int_{g_{n_1,\ldots,n_d}}^{g_{n_1+1,\ldots,n_d+1}} dF(\mathbf{v}) + P(g_{n_1,\ldots,n_d} \le \mathbf{X} \le g_{n_1+1,\ldots,n_d+1}) \\ &\le \left(\int_{a_{1,n_1}-h_1}^{a_{1,n_1}} + \int_{a_{1,n_1}}^{a_{1,n_1+1}} + \int_{a_{1,n_1+1}}^{a_{1,n_1+1}+h_1}\right)\cdots\left(\int_{a_{d,n_d}-h_1}^{a_{d,n_d}} + \int_{a_{d,n_d}}^{a_{d,n_d+1}} + \int_{a_{d,n_d+1}}^{a_{d,n_d+1}+h_1}\right) dF(\mathbf{v}) \\ &\quad - \int_{g_{n_1,\ldots,n_d}}^{g_{n_1+1,\ldots,n_d+1}} dF(\mathbf{v}) + \frac{1}{n}. \end{aligned}$$

Within this sum, the $3^d - 2^d$ terms with $\int_{a_{\alpha,n_\alpha}}^{a_{\alpha,n_\alpha+1}}$ are $O(n^{-1})$, while each of the 2^d terms without $\int_{a_{\alpha,n_\alpha}}^{a_{\alpha,n_\alpha+1}}$ is bounded by $h_{\text{prod}} \max_{\mathbf{x}\in R^d} |f(\mathbf{x})|$. Applying Assumptions (A1) and (A3),

$$E\int_{g_{n_1,\ldots,n_d}}^{\mathbf{x}} |K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|d\mathbf{u} \le Ch_{\text{prod}} \max_{\mathbf{x}\in R^d} |f(\mathbf{x})| + \frac{C(3^d - 2^d)}{n} = u(n^{-1/2}(\log n)^{1/2}). \qquad \blacksquare$$

LEMMA A8 *Under the same conditions of Lemma A7, for* $\forall \mathbf{x} = (x_1,\ldots,x_d) \in R^d$, $n^{-1}\sum_{i=1}^{n} |\zeta_{in}| = U_{a.s.}(n^{-1/2}\log n)$ *in which*

$$\zeta_{in} = \zeta_{i,n}(g_{n_1,\ldots,n_d}) = \int_{g_{n_1,\ldots,n_d}}^{\mathbf{x}} \{|K_{\mathbf{h}}(\mathbf{X}_i-\mathbf{u})|d\mathbf{u} - E|K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|\}d\mathbf{u},$$

while for i.i.d. $\mathbf{X}_1,\ldots,\mathbf{X}_n$, $n^{-1}\sum_{i=1}^{n} |\zeta_{in}| = U_{a.s.}(n^{-1/2}(\log n)^{1/2})$.

Proof One can show by applying Lemma A2 as in the proof of Lemma A5. $\blacksquare$

Proof of Theorem 3 Under the same conditions of Lemma A7, one has

$$\max_{1\le n_\alpha \le N_\alpha} |\hat{F}(g_{n_1,\ldots,n_d}) - F(g_{n_1,\ldots,n_d})| = O_{a.s.}(n^{-1/2}\log n)$$

by Lemma A5. For $\forall \mathbf{x} = (x_1,\ldots,x_d) \in R^d$, there exist integers $n_1,\ldots,n_d$ such that $F(g_{n_1,\ldots,n_d}) \le F(\mathbf{x}) \le F(g_{n_1+1,\ldots,n_d+1})$. Hence $|\hat{F}(\mathbf{x}) - \hat{F}(g_{n_1,\ldots,n_d})|$ is bounded by

$$\begin{aligned} &\left|\frac{1}{n}\sum_{i=1}^{n}\int_{g_{n_1,\ldots,n_d}}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i-\mathbf{u})d\mathbf{u}\right| \le \frac{1}{n}\sum_{i=1}^{n}\int_{g_{n_1,\ldots,n_d}}^{\mathbf{x}} |K_{\mathbf{h}}(\mathbf{X}_i-\mathbf{u})|d\mathbf{u} \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{g_{n_1,\ldots,n_d}}^{\mathbf{x}} \{|K_{\mathbf{h}}(\mathbf{X}_i-\mathbf{u})|d\mathbf{u} - E|K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|\}d\mathbf{u} + \int_{g_{n_1,\ldots,n_d}}^{\mathbf{x}} E|K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|d\mathbf{u} = O_{a.s.}(n^{-1/2}\log n) \end{aligned}$$

according to Lemmas A7 and A8. Then according to Lemma A5,

$$|\hat{F}(\mathbf{x}) - F(\mathbf{x})| \leq |\hat{F}(\mathbf{x}) - \hat{F}(g_{n_1,\ldots,n_d})| + |\hat{F}(g_{n_1,\ldots,n_d}) - F(g_{n_1,\ldots,n_d})| + |F(g_{n_1,\ldots,n_d}) - F(\mathbf{x})|$$
$$= U_{a.s.}(n^{-1/2}\log n) + U_{a.s.}(n^{-1/2}\log n) + U\left(\frac{1}{n}\right)$$

and if $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d, one can replace $\log n$ in this inequality by $(\log n)^{1/2}$. ■