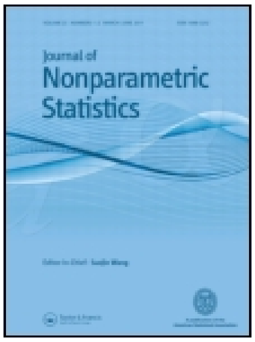


Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

Smooth simultaneous confidence bands for cumulative distribution functions

Jiangyan Wang , Fuxia Cheng & Lijian Yang

To cite this article: Jiangyan Wang , Fuxia Cheng & Lijian Yang (2013) Smooth simultaneous confidence bands for cumulative distribution functions, Journal of Nonparametric Statistics, 25:2, 395-407, DOI: 10.1080/10485252.2012.759219

To link to this article: https://doi.org/10.1080/10485252.2012.759219

Journal of Nonparametric Statistics, 2013
Vol. 25, No. 2, 395–407, http://dx.doi.org/10.1080/10485252.2012.759219

Smooth simultaneous confidence bands for cumulative distribution functions

Jiangyan Wang[a], Fuxia Cheng[b] and Lijian Yang[a,c]*

[a]*Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou 215006, China;* [b]*Department of Mathematics, Illinois State University, Normal, IL 61790, USA;* [c]*Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA*

(*Received 23 July 2012; final version received 7 December 2012*)

A plug-in kernel estimator is proposed for Hölder continuous cumulative distribution function (cdf) based on a random sample. Uniform closeness between the proposed estimator and the empirical cdf estimator is established, while the proposed estimator is smooth instead of a step function. A smooth simultaneous confidence band is constructed based on the smooth distribution estimator and the Kolmogorov distribution. Extensive simulation study using two different automatic bandwidths confirms the theoretical findings.

Keywords: bandwidth; confidence band; Hölder continuity; kernel; Kolmogorov distribution

AMS Subject Classifications: 62G08; 62G10; 62G20

1. Introduction

Consider a random sample $X_1, X_2, \ldots, X_n$ with a common cumulative distribution function (cdf) F. A well-known estimator of F is the empirical cdf

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

which possesses many desirable traits, among which are its invariance and the accompanying simultaneous confidence band for F. One serious drawback of F_n is its discontinuity, regardless of F being continuous or discrete.

To remedy this deficiency of F_n, Yamato (1973) proposed the following kernel distribution estimator:

$$\hat{F}(x) = \int_{-\infty}^{x} n^{-1} \sum_{i=1}^{n} K_h(u - X_i)\, \mathrm{d}u, \quad u \in \mathbb{R}, \tag{2}$$

in which $h = h_n > 0$ is called the bandwidth, K is a continuous probability density function (pdf) called kernel and $K_h(u) = K(u/h)/h$. The motivation of estimator $\hat{F}(x)$ is as follows. If $F(x) = \int_{-\infty}^{x} f(u)\, \mathrm{d}u$ for a pdf f, then $\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(u)\, \mathrm{d}u$ in which $\hat{f}(u) = n^{-1} \sum_{i=1}^{n} K_h(u - X_i)$

*Corresponding author. Email: yanglijian@suda.edu.cn, yangli@stt.msu.edu

Table 1. Critical values of the Kolmogorov–Smirnov goodness-of-fit test.

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
30	0.290	0.240	0.220	0.190
50	0.230	0.190	0.170	0.150
>50	$1.63/\sqrt{n}$	$1.36/\sqrt{n}$	$1.22/\sqrt{n}$	$1.07/\sqrt{n}$

is the well-known kernel density estimator of $f(u)$. One therefore can regard $\hat{F}$ as a plug-in integration estimator of F, which is always continuous. This line of research was followed up by Reiss (1981) and Falk (1985), and more recently extended to multivariate distribution in Liu and Yang (2008). Other related works on smooth estimation of cdf include Cheng and Peng (2002) and Xue and Wang (2010). What have been established are the asymptotic normal distribution of $\hat{F}$ identical to that of F_n and uniform rate of convergence of $\hat{F}$ to F. Yet, there do not exist any results on the closeness of $\hat{F}$ to F_n.

Denote the maximal deviation of F_n as

$$D_n(F_n) = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)|. \tag{3}$$

Then, one has the following classic result:

$$P\{\sqrt{n}D_n(F_n) \le \lambda\} \longrightarrow L(\lambda), \quad \text{as } n \longrightarrow \infty, \tag{4}$$

where $L(\lambda)$ is the well-known Kolmogorov distribution function, defined as

$$L(\lambda) \equiv 1 - 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp(-2j^2\lambda^2), \quad \lambda > 0. \tag{5}$$

Table 1 displays the percentiles of $D_n(F_n)$, which are critical values λ for the two-sided Kolmogorov–Smirnov test. In particular, for $n > 50$, the $100(1-\alpha)$th percentile is simply $L_{1-\alpha}/\sqrt{n}$, where $L_{1-\alpha} = L^{-1}(1-\alpha)$.

According to Equation (4), an asymptotic simultaneous confidence band can be computed for F as $F_n(x) \pm L_{1-\alpha}/\sqrt{n}, x \in \mathbb{R}$. More precisely, the confidence band is $[\max(F_n(x) - L_{1-\alpha}/\sqrt{n}, 0), \min(F_n(x) + L_{1-\alpha}/\sqrt{n}, 1)],\ x \in \mathbb{R}$.

Simultaneous confidence bands are powerful tools for global inference of complex functions, see, for instance, the confidence band for pdf in Bickel and Rosenblatt (1973) and for regression function in Wang and Yang (2009). To the best of our knowledge, there does not yet exist any smooth simultaneous confidence bands for F in the literature (the confidence band based on empirical cdf F_n is not smooth). In this paper, we show the uniform closeness between F_n and $\hat{F}$ to the order of $o_p(n^{-1/2})$, and therefore, a smooth simultaneous confidence band $[\max(\hat{F}(x) - L_{1-\alpha}/\sqrt{n}, 0), \min(\hat{F}(x) + L_{1-\alpha}/\sqrt{n}, 1)],\ x \in \mathbb{R}$, is obtained by replacing F_n with the smooth estimator $\hat{F}$. Furthermore, since $\hat{F}$ inherits all the asymptotic properties of F_n according to our Theorem 2.1, all existing results on the asymptotic normal distribution of $\hat{F}$ and its uniform weak convergence to F follow directly.

The rest of the paper is organised as follows. The main theoretical result on uniform asymptotics (Theorem 2.1) is given in Section 2. Data-driven implementation of procedures is described in Section 3, with simulation results presented in Section 4. All technical proofs are given in the appendix.

2. Main results

In this section, we establish that the two estimators $\hat{F}(x)$ and $F_n(x)$ are uniformly close under Hölder continuity. For nonnegative integer ν and $\delta \in (0, 1]$, denote by $C^{(\nu,\delta)}(\mathbb{R})$ the space of functions whose νth derivatives satisfy Hölder conditions of order δ, that is,

$$C^{(\nu,\delta)}(\mathbb{R}) = \left\{ \phi : \mathbb{R} \to \mathbb{R} \,\middle|\, \|\phi\|_{\nu,\delta} = \sup_{-\infty<x<y<+\infty} \frac{|\phi^{(\nu)}(x) - \phi^{(\nu)}(y)|}{|x-y|^{\delta}} < +\infty \right\}. \tag{6}$$

We state the following broad assumptions, for $\nu = 0, 1$ and some $\delta \in (0, 1]$.

(A1) *The cdf* $F \in C^{(\nu,\delta)}(\mathbb{R})$.
(A2) *The bandwidth* $h = h_n > 0$ *and* $\sqrt{n}h_n^{\nu+\delta} \to 0$.
(A3) *The kernel function* $K(\cdot)$ *is a continuous and symmetric probability density function, supported on* $[-1, 1]$.

THEOREM 2.1 *Under Assumptions* (A1)–(A3), *as* $n \to \infty$, *the maximal deviation between* $\hat{F}(x)$ *and* $F_n(x)$ *satisfies*

$$\sup_{x\in\mathbb{R}} |\hat{F}(x) - F_n(x)| = o_p(n^{-1/2}).$$

Denoting

$$D_n(\hat{F}) = D_n(\hat{F}, h) = \sup_{x\in\mathbb{R}} |\hat{F}(x) - F(x)| \tag{7}$$

and combining Equation (5) and Theorem 2.1, which provides that the maximal deviation between $\hat{F}(x)$ and $F_n(x)$ over the real line is of the order $o_p(n^{-1/2})$, one has the following corollary.

COROLLARY 2.2 *Under Assumptions* (A1)–(A3), *as* $n \to \infty$,

$$P\{\sqrt{n}D_n(\hat{F}) \le \lambda\} \to 1 - 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp(-2j^2\lambda^2), \quad \lambda > 0.$$

Hence, for any $\alpha \in (0, 1)$,

$$\lim_{n\to\infty} P\left\{F(x) \in \hat{F}(x) \pm \frac{L_{1-\alpha}}{\sqrt{n}}, x \in \mathbb{R}\right\} = 1 - \alpha,$$

and a smooth simultaneous confidence band for F is

$$\left[\max\left(\hat{F}(x) - \frac{L_{1-\alpha}}{\sqrt{n}}, 0\right), \min\left(\hat{F}(x) + \frac{L_{1-\alpha}}{\sqrt{n}}, 1\right)\right], \quad x \in \mathbb{R}.$$

We should point out that Theorem 2.1 holds without additional assumption to prevent h from being too small. This is because we estimate the distribution function $F(x)$ rather than a pdf. In fact, $\hat{F}(x)$ can be defined alternatively as in Equation (A2) which allows even $h = 0$, in which case $\hat{F}(x) \equiv F_n(x)$. This, of course, is not what one would prefer, as $h > 0$ yields smooth $\hat{F}(x)$. If a pdf $f(x) = F'(x)$ exists, a different bandwidth h should be used for estimating $f(x)$ by kernel smoothing, the optimal order of which is $n^{-1/5}$ if $F \in C^3(\mathbb{R})$.

In addition to the maximal deviation, the mean integrated squared error (MISE) is another useful measure of the performance of $\hat{F}$ and F_n. They are defined as

$$\mathrm{MISE}(\hat{F}) = \mathrm{MISE}(\hat{F}, h) = E \int \{\hat{F}(x) - F(x)\}^2 \, \mathrm{d}F(x), \tag{8}$$

$$\mathrm{MISE}(F_n) = \mathrm{MISE}(F_n, h) = E \int \{F_n(x) - F(x)\}^2 \, \mathrm{d}F(x). \tag{9}$$

Denote in what follows $G(x) = \int_{-\infty}^{x} K(u)\,\mathrm{d}u$, $\mu_2(K) = \int K(u)u^2\,\mathrm{d}u > 0$, and $D(K) = 1 - \int_{-1}^{1} G^2(u)\,\mathrm{d}u > 0$. According to Theorem 2 of Liu and Yang (2008),

$$\mathrm{MISE}(\hat{F}, h) = \mathrm{AMISE}(\hat{F}, h) + o(n^{-1}h + h^4),$$
$$\mathrm{AMISE}(\hat{F}, h) = \mathrm{MISE}(F_n) + 4^{-1}\mu_2^2(K)h^4 \int F''(x)^2 F'(x)\,\mathrm{d}x - n^{-1}D(K)h \int F'(x)^2\,\mathrm{d}x,$$

under the additional assumptions of $F \in C^3(\mathbb{R})$ and $nh \to \infty$. Elementary calculus shows that the optimal bandwidth (see also Yang and Tschernig 1999)

$$h_{\mathrm{opt}} = \left\{ \frac{1}{n} \frac{D(K)}{\mu_2^2(K)} \frac{C(F)}{B(F)} \right\}^{1/3} \tag{10}$$

with $B(F) = \int F''(x)^2 F'(x)\,\mathrm{d}x$, $C(F) = \int F'(x)^2\,\mathrm{d}x$, which minimises the asymptotic mean integrated squared error (AMISE) and the corresponding minimum

$$\mathrm{AMISE}(\hat{F}, h_{\mathrm{opt}}) = \mathrm{MISE}(F_n) - \frac{3}{4} n^{-4/3} \left\{ \frac{D^4(K)C^4(F)}{\mu_2^2(K)B(F)} \right\}^{1/3} < \mathrm{MISE}(F_n).$$

Therefore, the estimator $\hat{F}$ with bandwidth h_{opt} has smaller AMISE than the empirical cdf F_n, although this advantage diminishes with $n \to \infty$ due to the order $n^{-4/3}$ smaller than the order of $\mathrm{MISE}(F_n) = n^{-1} \int F(x)\{1 - F(x)\}\,\mathrm{d}F(x)$. We agree with one referee that choosing an optimal bandwidth in terms of coverage probability of the smooth confidence band remains open as the dependence of coverage probability on a bandwidth is extremely complicated.

3. Implementation

In this section, we describe the procedures to construct the smooth confidence band $[\max(\hat{F}(x) - L_{1-\alpha}/\sqrt{n}, 0), \min(\hat{F}(x) + L_{1-\alpha}/\sqrt{n}, 1)]$, $x \in \mathbb{R}$.

To begin with, one computes $L_{1-\alpha}$ in Equation (4) using Table 1 and $\hat{F}(x)$ according to Equation (2) using the quartic kernel $K(u) = 15(1 - u^2)^2 I\{|u| \le 1\}/16$ and two candidate bandwidths h_1 and h_2. The bandwidth $h_1 = \mathrm{IQR} \times n^{-1}$ in which IQR is the sample inter-quartile range of $\{X_i\}_{i=1}^n$ and h_2 is a modified version of the plug-in optimal bandwidth of Liu and Yang (2008). To be precise

$$h_2 = \hat{h}_{\mathrm{opt}} = n^{-1/3} \left\{ \frac{D(K)}{\mu_2^2(K)} \right\}^{1/3} \left\{ \frac{\hat{C}(F)}{\hat{B}(F)} \right\}^{1/3}, \tag{11}$$

where $\hat{B}(F)$ and $\hat{C}(F)$ are plug-in estimators of $B(F)$ and $C(F)$, respectively,

$$\hat{B}(F) = n^{-1} \sum_{j=1}^{n} \left\{ n^{-1} \sum_{i=1}^{n} K'_{h_{\mathrm{rot}}}(X_j - X_i) \int_{-\infty}^{X_j} K_{h_{\mathrm{rot}}}(x - X_i)\,\mathrm{d}x \right\}^2,$$

$$\hat{C}(F) = n^{-1} \sum_{j=1}^{n} \left\{ n^{-1} \sum_{i=1}^{n} K_{h_{\mathrm{rot}}}(X_j - X_i) \int_{-\infty}^{X_j} K_{h_{\mathrm{rot}}}(x - X_i)\,\mathrm{d}x \right\},$$

with the rule-of-thumb pilot bandwidth $h_{\mathrm{rot}} = 2.78 \times (\mathrm{IQR}/1.349) \times n^{-1/5}$.

Equations (10) and (11) show that the bandwidth h_2 is AMISE optimal under the strong assumption of $F \in C^{(3)}(\mathbb{R})$. In contrast, h_1 satisfies Assumption (A2) as long as F satisfies the much weaker Hölder condition of order $\delta > 1/2$.

4. Simulation

In this section, we compare the global performance of the two estimators F_n and $\hat{F}$, in terms of their errors and the simultaneous confidence bands for F. The global performance of F_n and $\hat{F}$ can be measured in terms of the maximal deviation $D_n(F_n)$, $D_n(\hat{F})$, $\mathrm{MISE}(F_n)$ and $\mathrm{MISE}(\hat{F})$, respectively, as defined in Equations (3), (7), (9) and (8).

4.1. *Global performance of the two estimators*

In this section, we examine the asymptotic results of Equations (3), (7), (8) and (9) via simulation experiments, using the quartic kernel and the two bandwidths described in Section 3. The data set $\{X_i\}_{i=1}^n$ is generated from the standard normal distribution, the standard exponential distribution and the standard Cauchy distribution, respectively. In other words, $F(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \mathrm{e}^{-u^2/2}\,\mathrm{d}u$, $F(x) = (1 - \mathrm{e}^{-x}) I(x > 0)$ or $F(x) = \pi^{-1} \arctan x + 1/2$.

The quantities $D_n(F_n), D_n(\hat{F}), \mathrm{MISE}(\hat{F})$ and $\mathrm{MISE}(F_n)$ are computed according to Equations (3), (7), (8) and (9), respectively, with sample size $n = 100, 200, 500, 1000$, by running over 1000 replications for bandwidths h_1 and h_2. Of interest are the means over the 1000 replications of $D_n(\hat{F})$ and $D_n(F_n)$ defined in Equations (7) and (3), denoting as $\bar{D}_n(\hat{F})$ and $\bar{D}_n(F_n)$, respectively. It is similar for $\mathrm{MISE}(\hat{F})$ and $\mathrm{MISE}(F_n)$. Both measures are listed in Table 2, which contains the means of all the ratios, using both bandwidths h_1 and h_2.

Table 2. $\bar{D}_n(\hat{F})/\bar{D}_n(F_n)$ and $\mathrm{MISE}(\hat{F})/\mathrm{MISE}(F_n)$ from 1000 replicates and two bandwidths for the standard normal, exponential and Cauchy distribution, respectively (from left to right).

	$\bar{D}_n(\hat{F})/\bar{D}_n(F_n)$			$\mathrm{MISE}(\hat{F})/\mathrm{MISE}(F_n)$		
n	Normal	Exponential	Cauchy	Normal	Exponential	Cauchy
100	0.983	0.979	0.981	0.977	0.976	0.973
	0.674	0.991	0.758	0.811	1.002	0.870
200	0.989	0.985	0.988	0.990	0.982	0.986
	0.698	1.103	0.773	0.845	1.033	0.897
500	0.993	0.991	0.993	0.996	0.991	0.994
	0.730	1.320	0.791	0.881	1.124	0.928
1000	0.995	0.993	0.995	0.999	0.995	0.995
	0.751	1.495	0.799	0.905	1.171	0.937

Note: The numbers above/below are the ratios by using h_1 and h_2, respectively.

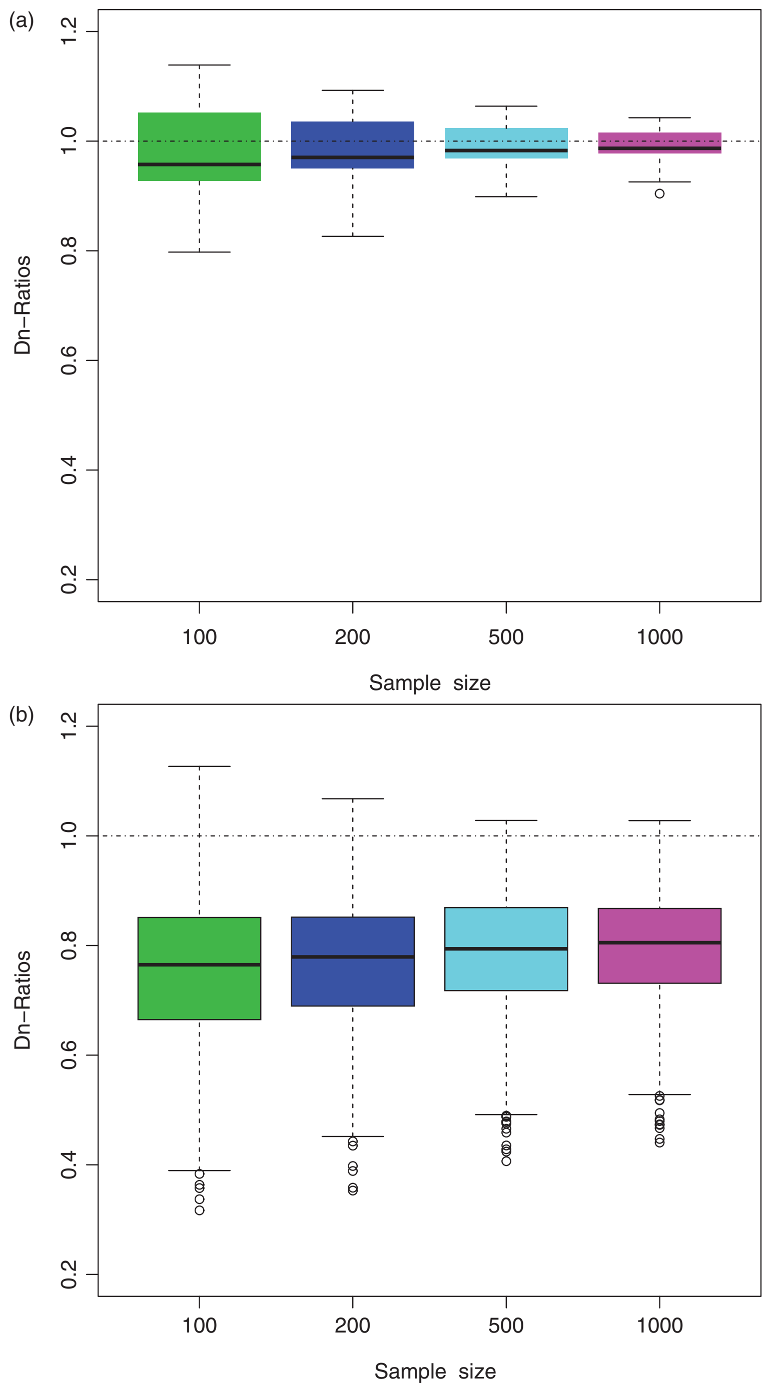


Figure 1. The ratios of $D_n(\hat{F})/D_n(F_n)$ for the standard normal distribution by using two bandwidths over 1000 replications. The bandwidth of (a) is h_1 and (b) is h_2.

Boxplot of all the ratios is displayed in Figures 1–3 for the standard normal distribution, the standard exponential distribution and the standard Cauchy distribution. These figures clearly show that for all of the distributions, the smooth estimator $\hat{F}$ with bandwidth h_1 and the empirical cdf

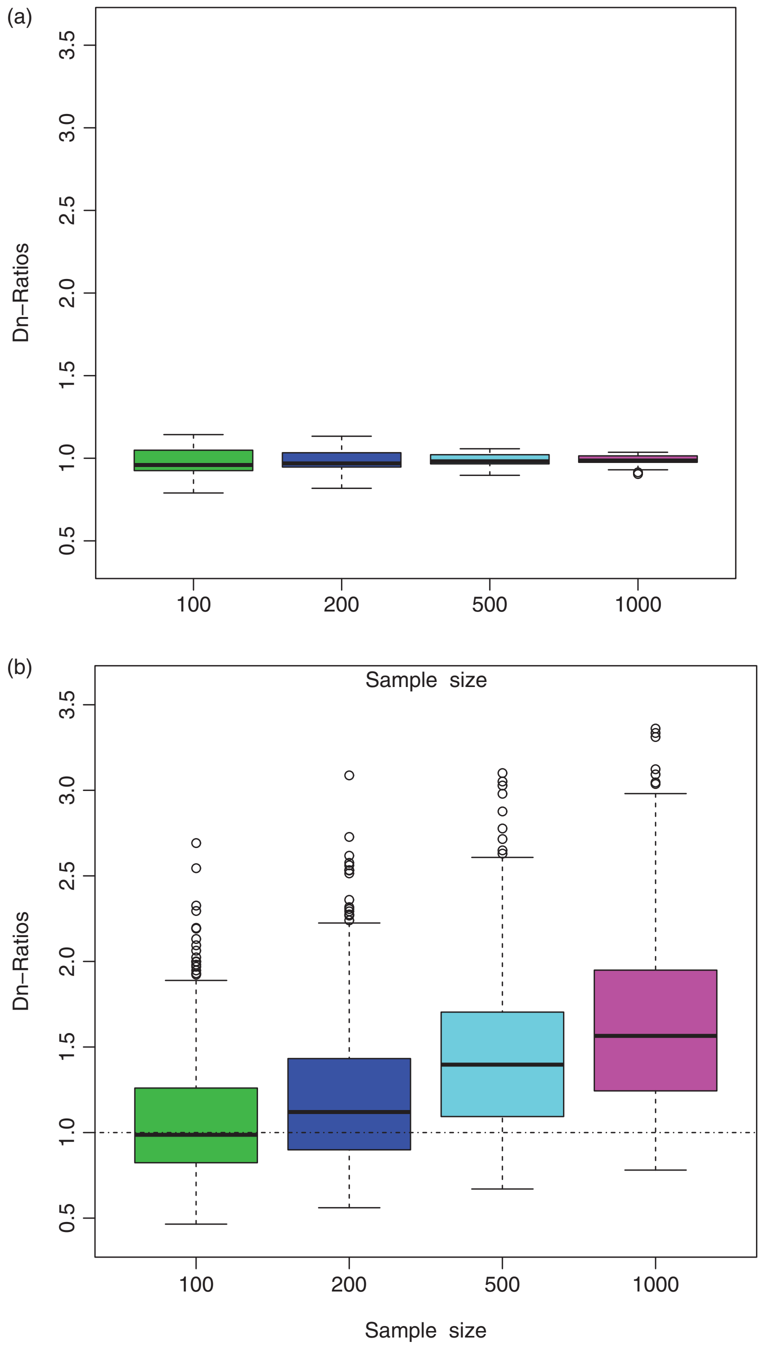
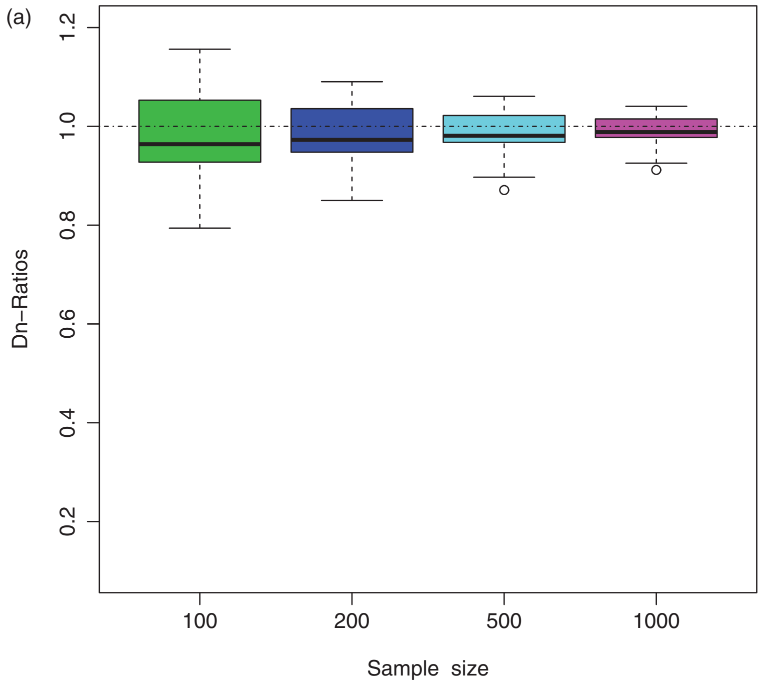


Figure 2. The ratios of $D_n(\hat{F})/D_n(F_n)$ for the standard exponential distribution by using two bandwidths over 1000 replications. The bandwidth of (a) is h_1 and (b) is h_2.

F_n become close rapidly (i.e. the ratio $D_n(\hat{F})/D_n(F_n)$ converges to 1 in probability), which is consistent with the result of Theorem 2.1. For the standard normal distribution and the standard Cauchy distribution, the smooth estimator $\hat{F}$ with bandwidth h_2 is much more efficient than the empirical estimator F_n, especially for smaller sample sizes, as the bandwidth h_2 is AMISE



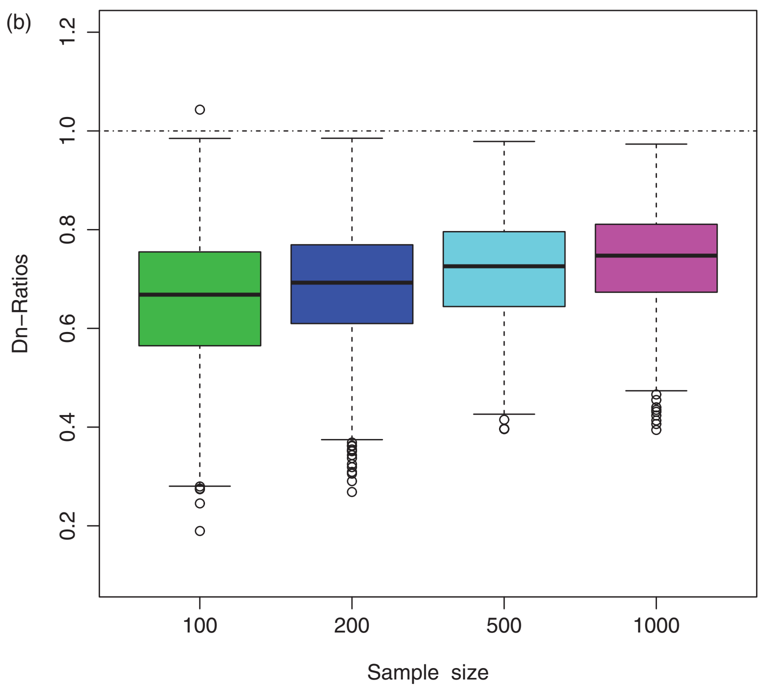


Figure 3. The ratios of $D_n(\hat{F})/D_n(F_n)$ for the standard Cauchy distribution by using two bandwidths over 1000 replications. The bandwidth of (a) is h_1 and (b) is h_2.

optimal when the true cdf $F \in C^{(3)}(\mathbb{R})$. For the standard exponential distribution, on the other hand, the $\hat{F}$ with bandwidth h_2 is much less efficient than the empirical estimator F_n, due to the true cdf $F(x) = (1 - \mathrm{e}^{-x})I(x > 0) \notin C^{(3)}(\mathbb{R})$ (see the discussion at the end of Section 3). The same phenomenon is observed for MISE($\hat{F}$) and MISE(F_n), but graphs are omitted. Table 2 is the numerical summary of the above graphical observations.

4.2. *Simultaneous confidence bands*

In this section, we compare by simulations the behaviour of the proposed smooth confidence bands for the same data set $\{X_i\}_{i=1}^n$ generated from the standard normal distribution, the standard exponential distribution and the standard Cauchy distribution, respectively.

We take the confidence level $1-\alpha = 0.99$, 0.95, 0.90, 0.80, respectively, and sample size $n = 30$, 50, 100, 200, 500, 1000, respectively. Tables 3–5 contain the frequencies over 1000 replications of coverage at all data points $\{X_i\}_{i=1}^n$, for two types of confidence bands by using $\hat{F}$ and F_n, respectively.

Figures 4–6 depict the true cdf F (thick), the $\hat{F}$ together with its smooth 90% confidence band (solid) and F_n (dashed) for the normal, exponential and Cauchy distributions, respectively. The samples used has size $n = 100$, and bandwidth h_1 is used for computing $\hat{F}$ in (a) and h_2 in (b). Similar patterns have been observed for samples of size $n = 500$, but not included due to space constraint.

Table 3. Coverage frequencies for the standard normal distribution from 1000 replications.

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
30	0.995 (0.994) 0.999	0.973 (0.961) 0.998	0.946 (0.945) 0.992	0.867 (0.857) 0.977
50	0.993 (0.993) 0.999	0.976 (0.965) 0.996	0.943 (0.924) 0.985	0.868 (0.873) 0.968
100	0.994 (0.991) 0.999	0.965 (0.963) 0.992	0.938 (0.930) 0.987	0.858 (0.851) 0.965
200	0.987 (0.989) 0.998	0.955 (0.954) 0.990	0.919 (0.914) 0.981	0.844 (0.838) 0.953
500	0.993 (0.991) 0.999	0.958 (0.960) 0.990	0.922 (0.917) 0.978	0.839 (0.831) 0.950
1000	0.990 (0.990) 0.998	0.960 (0.959) 0.989	0.909 (0.908) 0.975	0.820 (0.816) 0.944

Notes: The numbers outside of the parentheses are the confidence band coverage frequencies of $\hat{F}$ by using the bandwidths h_1 (left) and h_2 (right), respectively, and inside are the coverage frequencies of F_n.

Table 4. Coverage frequencies for the standard Cauchy distribution from 1000 replications.

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
30	0.996 (0.995) 1.000	0.978 (0.972) 0.997	0.958 (0.946) 0.990	0.889 (0.862) 0.959
50	0.997 (0.994) 1.000	0.976 (0.966) 0.993	0.932 (0.923) 0.982	0.860 (0.841) 0.953
100	0.993 (0.991) 1.000	0.964 (0.963) 0.991	0.932 (0.918) 0.980	0.859 (0.845) 0.952
200	0.992 (0.993) 1.000	0.966 (0.963) 0.992	0.912 (0.908) 0.980	0.834 (0.825) 0.933
500	0.994 (0.991) 0.998	0.958 (0.956) 0.988	0.902 (0.894) 0.965	0.806 (0.803) 0.922
1000	0.996 (0.996) 1.000	0.969 (0.967) 0.992	0.927 (0.918) 0.980	0.822 (0.816) 0.940

Notes: The numbers outside of the parentheses are the confidence band coverage frequencies of $\hat{F}$ by using the bandwidths h_1 (left) and h_2 (right), respectively, and inside are the coverage frequencies of F_n.

Table 5. Coverage frequencies for the standard exponential distribution from 1000 replications.

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
30	0.996 (0.996) 1.000	0.977 (0.973) 0.998	0.956 (0.945) 0.994	0.890 (0.864) 0.977
50	0.994 (0.995) 0.999	0.977 (0.970) 0.994	0.948 (0.928) 0.991	0.886 (0.866) 0.973
100	0.996 (0.997) 1.000	0.974 (0.969) 0.994	0.921 (0.914) 0.991	0.843 (0.845) 0.938
200	0.994 (0.993) 0.999	0.955 (0.953) 0.989	0.909 (0.908) 0.964	0.817 (0.799) 0.829
500	0.994 (0.994) 1.000	0.976 (0.959) 0.816	0.920 (0.919) 0.845	0.838 (0.835) 0.450
1000	0.990 (0.989) 0.996	0.963 (0.962) 0.832	0.923 (0.911) 0.390	0.829 (0.829) 0.050

Notes: The numbers outside of the parentheses are the confidence band coverage frequencies of $\hat{F}$ by using the bandwidths h_1 (left) and h_2 (right), respectively, and inside are the coverage frequencies of F_n.

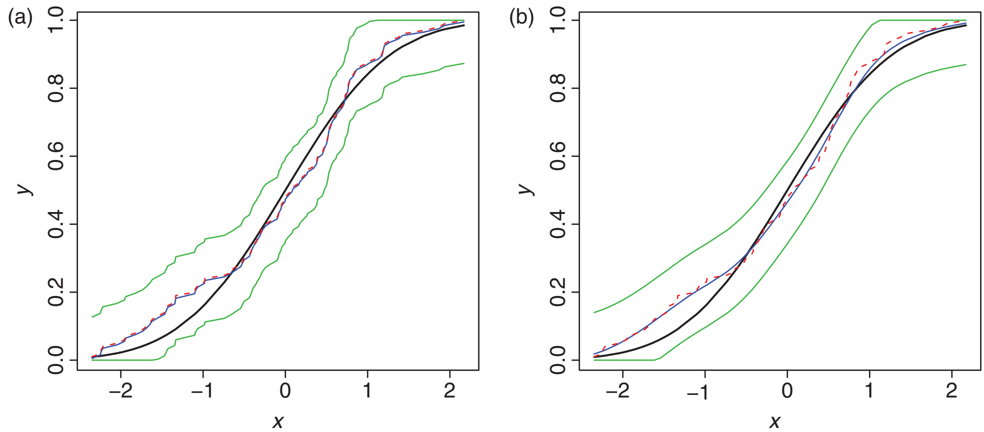


Figure 4. Smooth simultaneous confidence bands for the standard normal distribution at $\alpha = 0.1$ and $n = 100$. The bandwidth of (a) and (b) is h_1 and h_2, respectively.

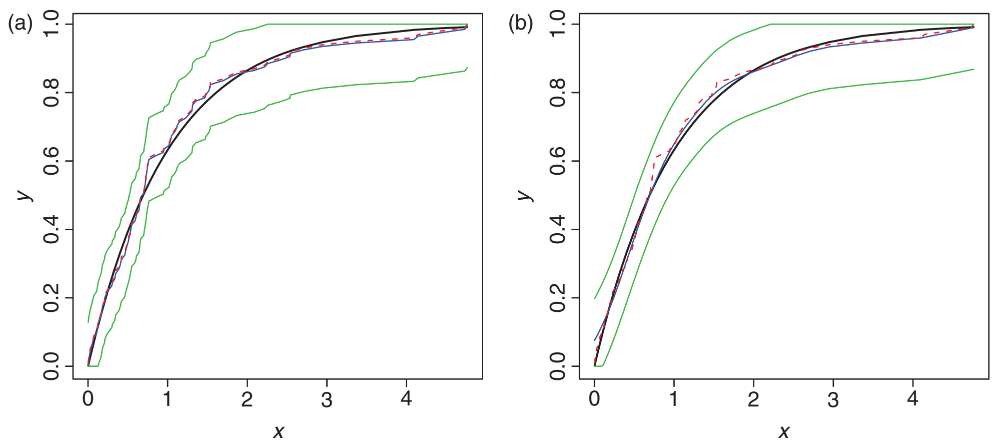


Figure 5. Smooth simultaneous confidence bands for the standard exponential distribution at $\alpha = 0.1$ and $n = 100$. The bandwidth of (a) and (b) is h_1 and h_2, respectively.

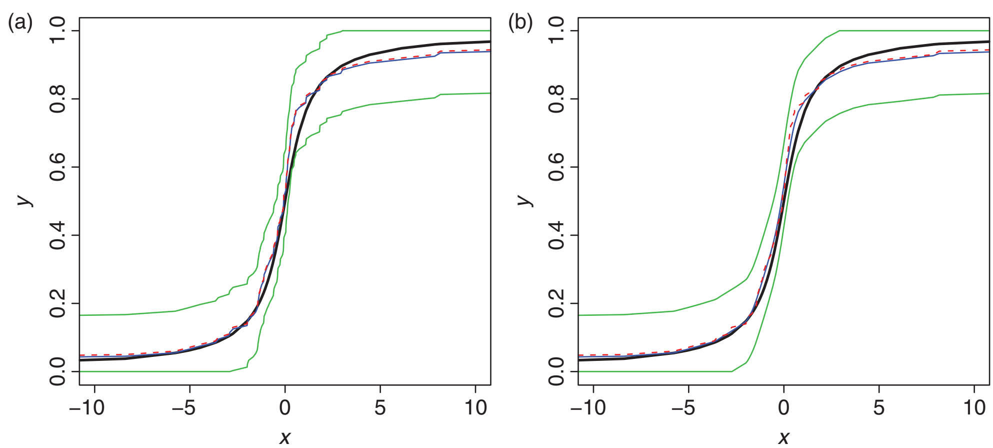


Figure 6. Smooth simultaneous confidence bands for the standard Cauchy distribution at $\alpha = 0.1$ and $n = 100$. The bandwidth of (a) and (b) is h_1 and h_2, respectively.

For all three distributions, it is obvious that the smooth confidence band with bandwidth h_1 has almost the same coverage frequencies as the unsmooth confidence band based on F_n. While both are conservative, increasing sample size does bring coverage frequencies closer to the nominal coverage probabilities.

The confidence band using h_2 has coverage frequencies higher than the nominal for normal and Cauchy distributions, and unacceptably below the nominal for the exponential distribution. The latter is due to the fact that $F \in C^{(0,1)}(\mathbb{R})$, $\nu + \delta = 1$ for the exponential distribution and so Assumption (A2) does not hold for $h_2 \sim n^{-1/3}$.

To summarise, in terms of estimation error, one should use bandwidth h_1 as a robust default to compute $\hat{F}$. Bandwidth h_2 would be highly recommended if the true cdf F was known to be third-order smooth, in which case both the maximal deviation $D_n(\hat{F})$ and the mean integrated squared error $\mathrm{MISE}(\hat{F})$ could be substantially reduced. In terms of coverage probabilities of the smooth simultaneous confidence band proposed in this paper, however, the bandwidth h_1 is always recommended over the bandwidth h_2.

Acknowledgements

This research has been supported in part by NSF Award DMS 1007594, Jiangsu Specially-Appointed Professor Program and Jiangsu Key Discipline Program (Statistics), Jiangsu, China. The helpful comments by the Editor-in-Chief and two referees are gratefully acknowledged.

References

Bickel, P., and Rosenblatt, M. (1973), 'On Some Global Measures of the Deviations of Density Function Estimates', *Annals of Statistics*, 1, 1071–1095.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

Cheng, M., and Peng, L. (2002), 'Regression Modeling for Nonparametric Estimation of Distribution and Quantile Functions', *Statistica Sinica*, 12, 1043–1060.

Falk, M. (1985), 'Asymptotic Normality of Kernel Type Estimators of Quantiles', *Annals of Statistics*, 13, 428–433.

Liu, R., and Yang, L. (2008), 'Kernel Estimation of Multivariate Cumulative Distribution Function', *Journal of Nonparametric Statistics*, 20, 661–667.

Reiss, R. (1981), 'Nonparametric Estimation of Smooth Distribution Functions', *Scandinavian Journal of Statistics. Theory and Applications*, 8, 116–119.

Wang, J., and Yang, L. (2009), 'Polynomial Spline Confidence Bands for Regression Curves', *Statistica Sinica*, 19, 325–342.

Xue, L., and Wang, J. (2010), 'Distribution Function Estimation by Constrained Polynomial Spline Regression', *Journal of Nonparametric Statistics*, 22, 443–457.

Yamato, H. (1973), 'Uniform Convergence of an Estimator of a Distribution Function', *Bulletin of Mathematical Statistics*, 15, 69–78.

Yang, L., and Tschernig, R. (1999), 'Multivariate Bandwidth Selection for Local Linear Regression', *Journal of Royal Statistical Society, Series B*, 61, 793–815.

Appendix

Throughout this section, we denote by c any positive constant and by u_p sequence of random functions of real values x and v which are o_p of certain order uniformly over $x \in \mathbb{R}$ and $v \in [-1, 1]$.

A.1. Proof of Theorem 2.1

Proof Recall the notation $G(x) = \int_{-\infty}^{x} K(u)\,\mathrm{d}u$; by Equation (2), one has

$$\hat{F}(x) = n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{x} K_h(u - X_i)\,\mathrm{d}u = n^{-1} \sum_{i=1}^{n} G\left(\frac{x - X_i}{h}\right). \tag{A1}$$

According to the definition of $F_n(x)$ in Equation (1), with integration by parts and a change of variable $v = (x-u)/h$, it follows that

$$\begin{aligned}\hat{F}(x) &= \int_{-\infty}^{+\infty} G\left(\frac{x-u}{h}\right) \mathrm{d}F_n(u) \\ &= G\left(\frac{x-u}{h}\right) F_n(u) \mid_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} F_n(u) K\left(\frac{x-u}{h}\right) \frac{1}{h} \,\mathrm{d}u \\ &= \int_{-\infty}^{+\infty} F_n(u) \frac{1}{h} K\left(\frac{x-u}{h}\right) \mathrm{d}u \\ &= \int_{-\infty}^{+\infty} F_n(x-hv) K(v) \,\mathrm{d}v. \end{aligned} \tag{A2}$$

Hence, one obtains that

$$\begin{aligned}\hat{F}(x) - F_n(x) &= \int_{-\infty}^{+\infty} \{F_n(x-hv) - F_n(x)\} K(v) \,\mathrm{d}v \\ &= \int_{-1}^{1} \{F_n(x-hv) - F_n(x)\} K(v) \,\mathrm{d}v. \end{aligned} \tag{A3}$$

Since $n \to \infty, \sqrt{n}\{F_n(x) - F(x)\} \to_D B(F(x))$, where B denotes the Brownian bridge; hence, the following holds (Billingsley 1968):

$$\sup_{v \in [-1,1]} \sup_{x \in \mathbb{R}} |\sqrt{n}\{F_n(x-hv) - F_n(x)\} - \sqrt{n}\{F(x-hv) - F(x)\}| = o_p(1).$$

Thus, one obtains

$$|\{F_n(x-hv) - F_n(x)\} - \{F(x-hv) - F(x)\}| = u_p(n^{-1/2}). \tag{A4}$$

By the assumptions on the cdf F, we treat separately the cases of $\nu = 1$ and $\nu = 0$.

Case 1 $\nu = 1$: According to Assumption (A1), applying Taylor expansion to $F \in C^{(1,\delta)}(\mathbb{R})$, there exists $c > 0$ such that for $x \in \mathbb{R}, v \in [-1, 1]$,

$$|F(x-hv) - F(x) - f(x)(-hv)| \leq ch^{1+\delta}|v|^{1+\delta} \leq ch^{1+\delta}.$$

Thus,

$$\left| \int_{-1}^{1} \{F(x-hv) - F(x) - f(x)(-hv)\} K(v) \,\mathrm{d}v \right| \leq ch^{1+\delta} \int_{-1}^{1} K(v) \,\mathrm{d}v \leq ch^{1+\delta};$$

since $\int_{-1}^{1} vK(v) \,\mathrm{d}v = 0$, one obtains that

$$\left| \int_{-1}^{1} \{F(x-hv) - F(x)\} K(v) \,\mathrm{d}v \right| = \left| \int_{-1}^{1} \{F(x-hv) - F(x) - f(x)(-hv)\} K(v) \,\mathrm{d}v \right| \leq ch^{1+\delta};$$

hence applying Equation (A4) and Assumption (A2),

$$\begin{aligned}&\left| \int_{-1}^{1} \{F_n(x-hv) - F_n(x)\} K(v) \,\mathrm{d}v \right| \\ &\quad\leq \int_{-1}^{1} |\{F(x-hv) - F(x) - f(x)(-hv)\}| K(v) \,\mathrm{d}v + u_p(n^{-1/2}) \\ &\quad\leq ch^{1+\delta} + u_p(n^{-1/2}) = u_p(n^{-1/2}). \end{aligned} \tag{A5}$$

Case 2 $\nu = 0$: According to Assumption (A1), $F \in C^{(0,\delta)}(\mathbb{R})$ and there exists $c > 0$ such that for $x \in \mathbb{R}, v \in [-1, 1]$,

$$|F(x - hv) - F(x)| \leq ch^{\delta}|v|^{\delta} \leq ch^{\delta}.$$

Applying Equation (A4) and Assumption (A2), one has

$$\begin{aligned} &\left|\int_{-1}^{1}\{F_n(x - hv) - F_n(x)\}K(v)\,\mathrm{d}v\right| \\ &\leq \left|\int_{-1}^{1}\{F(x - hv) - F(x)\}K(v)\,\mathrm{d}v\right| + u_p(n^{-1/2}) \\ &\leq ch^{\delta} + u_p(n^{-1/2}) = u_p(n^{-1/2}). \end{aligned} \tag{A6}$$

Thus, applying Equations (A3)–(A5) or (A6), it follows that

$$\begin{aligned} \sup_{x\in\mathbb{R}}|\hat{F}(x) - F_n(x)| &= \sup_{x\in\mathbb{R}}\left|\int_{-1}^{1}\{F_n(x - hv) - F_n(x)\}K(v)\,\mathrm{d}v\right| \\ &= o_p(n^{-1/2}). \end{aligned}$$

Therefore, one completes the proof of Theorem 2.1. ∎