DOI 10.1007/s10958-016-3207-1
Journal of Mathematical Sciences, Vol. 220, No. 6, February, 2017

KERNEL ESTIMATION OF A CHARACTERISTIC FUNCTION

A. A. Abdushukurov[1] **and F. Sh. Norboev**[2]

This article deals with the study of asymptotic properties of a nonparametric kernel estimator of a characteristic function. Uniformly strong consistency of kernel estimator with fixed and expanding interval is established. The law of iterated logarithm type is also proved.

Let $X_1, \ldots, X_n$ be independent repeated sample of a random variable X with distribution function (d.f.) $F(x)$, $|x| \leqslant \infty$. It is well known [2] that the empirical distribution function (e.d.f.)

$$F_n^{\mathrm{e}}(x) = \frac{1}{n} \sum_{k=1}^{n} I(X_k \leqslant x), \quad |x| \leqslant \infty,$$

is a good nonparametric estimator of $F(x)$, where $I(A)$ is the indicator of an event A. However, if F is absolutely continuous with the density $f(x) = F'(x)$, then $F_n^{\mathrm{e}}(x)$ is undesirable as an estimator of F since $F_n^{\mathrm{e}}(x)$ does not have a density. In such cases it is natural to use the smoothed kernel estimator

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} K_n(x - X_k) = \int_{-\infty}^{\infty} K_n(x - y) dF_n^{\mathrm{e}}(y),$$

where $K_n(x) = K\left(\frac{x}{b_n}\right)$, $K(x)$ is a given absolutely continuous d.f. (a kernel) with the density $k(x) = K'(x)$, and the sequence of "window widths" $\{b_n,\ n \geqslant 1\}$ is such that $b_n \downarrow 0$ as $n \to \infty$. It is easy to see that the density of $F_n(x)$ equals

$$f_n(x) = \frac{1}{nb_n} \sum_{k=1}^{n} k\left(\frac{x - X_k}{b_n}\right) = \frac{1}{b_n} \int_{-\infty}^{\infty} k\left(\frac{x - y}{b_n}\right) dF_n^{\mathrm{e}}(y).$$

Let us note that the basic properties of kernel estimators F_n and f_n are described in [1] and [2], respectively.

The goal of the present paper is to study the kernel estimate

$$c_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x) \tag{1}$$

of characteristic function (c.f.)

$$c(t) = \mathbf{M} e^{itX} = \int_{-\infty}^{\infty} e^{itx} dF(x),$$

corresponding to d.f. F. Let $c_n^{\mathrm{e}}(t)$ be empirical c.f. (e.c.f.), obtained from $c(t)$ by substituting F_n^{e} instead of F. Profound and, in our opinion, the ultimate results on the properties of e.c.f. are obtained in [5].

[1] National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan, e-mail: a_abdushukurov@rambler.ru
[2] Gulistan State University, Guliston, Provincia de Sirdarin, Uzbekistan, e-mail: asi_96@mail.ru

Translated from *Statisticheskie Metody Otsenivaniya i Proverki Gipotez*, Vol. 19, pp. 26–33, 2006
1072-3374/17/2206-0643

The estimate (1) is first met in [4]. Analysis of available studies showed that the kernel estimate of c.f. has not yet been studied extensively. In the present paper we mostly study the consistency properties of the estimate (1).

Let the kernel $K(x)$ have compact support $[c, d]$, $-\infty < c < d < \infty$. For two d.f. F_1, F_2, and their c.f. c_1 and c_2 we introduce the following norms:

$$\|F_1 - F_2\|_\infty = \sup_{|x| \leqslant \infty} |F_1(x) - F_2(x)|,$$

$$\|c_1 - c_2\|_{T_1}^{T_2} = \sup_{T_1 \leqslant t \leqslant T_2} |c_1(t) - c_2(t)|,$$

$$\|c_1 - c_2\|_T = \sup_{|t| \leqslant T} |c_1(t) - c_2(t)|,$$

where T, T_1, and T_2 are some real numbers.

The following lemma will be used repeatedly in what follows.

Lemma 1. *For any two c.f.* $c_1(t)$ *and* $c_2(t)$ *the following inequality holds*:

$$\|c_1 - c_2\|_{T_1}^{T_2} \leqslant 2(1 + T \cdot M_n)\|F_1 - F_2\|_\infty + [1 - F_1(M_n) + F_1(-M_n)] + \\ + [1 - F_2(M_n) + F_2(-M_n)], \tag{2}$$

where $T = \max(|T_1|, |T_2|)$ *and* $\{M_n, n \geqslant 1\}$ *is a sequence of nonnegative numbers* (*possibly random*), *tending to* $+\infty$ *as* $n \to \infty$ *with a certain speed.*

Proof. The following relations hold:

$$\|c_1 - c_2\|_{T_1}^{T_2} = \left\| \int_{-\infty}^{\infty} e^{itx} d(F_1(x) - F_2(x)) \right\|_{T_1}^{T_2} \leqslant \left\| \int_{|x| \leqslant M_n} e^{itx} d(F_1(x) - F_2(x)) \right\|_{T_1}^{T_2} + \\ + \left\| \int_{|x| > M_n} e^{itx} d\left(F_1(x) - F_2(x)\right) \right\|_{T_1}^{T_2} = R_{1n} + R_{2n}. \tag{3}$$

Integrating by parts, we have

$$R_{1n} \leqslant \left\|e^{itx}\right\|_{T_1}^{T_2} \cdot \left\|F_1(x) - F_2(x)\right\|_{-M_n}^{M_n} + \left\| it \int_{-M_n}^{M_n} e^{itx} (F_1(x) - F_2(x)) dx \right\|_{T_1}^{T_2} \leqslant \\ \leqslant 2(1 + T \cdot M_n)\|F_1 - F_2\|_\infty. \tag{4}$$

On the other hand,

$$R_{2n} \leqslant \left\| \int_{|x| > M_n} e^{itx} dF_1(x) \right\|_{T_1}^{T_2} + \left\| \int_{|x| > M_n} e^{itx} dF_2(x) \right\|_{T_1}^{T_2} \leqslant \\ \leqslant \int_{|x| > M_n} d(F_1(x) + F_2(x)) = [1 - F_1(M_n) + F_1(-M_n)] + \\ + [1 - F_2(M_n) + F_2(-M_n)]. \tag{5}$$

Then (2) follows from (3)–(5). Lemma 1 is proved.

Remark 1. Let $F_1 = F_n$ and $F_2 = F_n^{\mathrm{e}}$. Then

$$1 - F_n(M_n) + F_n(-M_n) \leqslant \left|F_n(M_n) - F_n^{\mathrm{e}}(M_n)\right| + \left|F_n(-M_n) - F_n^{\mathrm{e}}(-M_n)\right| +$$

$$+1 - F_n^{\mathrm{e}}(M_n) + F_n^{\mathrm{e}}(-M_n) \leqslant 2\left\|F_n - F_n^{\mathrm{e}}\right\|_\infty + 1 - F_n^{\mathrm{e}}(M_n) + F_n^{\mathrm{e}}(-M_n).$$

Then in view of (2)

$$\left\|c_n - c_n^{\mathrm{e}}\right\|_{T_1}^{T_2} \leqslant 2(2 + T \cdot M_n)\left\|F_n - F_n^{\mathrm{e}}\right\|_\infty + [1 - F_n^{\mathrm{e}}(M_n) + F_n^{\mathrm{e}}(-M_n)]. \tag{6}$$

In [3] the following estimate is established:

$$\left\|F_n - F_n^{\mathrm{e}}\right\|_{T_1}^{T_2} \leqslant \|Q_n(x)\|_\infty + \left\|\int\limits_c^d (F(x - b_n z) - F(x))\, dK(x)\right\|_\infty, \tag{7}$$

where

$$\|Q_n(x)\|_\infty = \sup_{-\infty \leqslant x \leqslant \infty} \sup_{|t| \leqslant e(c,d) b_n} \left|F_n^{\mathrm{e}}(x+t) - F(x+t) - \left[F_n^{\mathrm{e}}(x) - F(x)\right]\right|$$

and $e(c,d) = \max\{|c|, |d|\}$.

From [3] we also have

$$\|Q_n(x)\|_\infty \overset{\text{a.s.}}{=} o\left(\left(\frac{b_n \cdot \log\, n}{n}\right)^{1/2}\right), \quad n \to \infty. \tag{8}$$

Remark 2. Let $F_1 = F_n^*$ and $F_2 = F$, where

$$F_n^*(x) = \int\limits_c^d F(x - b_n z) dK(z) = MF_n(x).$$

Then

$$\|F_n^* - F_n\|_\infty = \left\|\int\limits_c^d (F(x - b_n z) - F(x))\, dK(z)\right\|_\infty, \tag{9}$$

and

$$1 - F_n^*(M_n) + F_n^*(-M_n) \leqslant \int\limits_c^d |F(M_n - b_n z) - F(M_n)|\, dK(z) +$$

$$+ \int\limits_c^d |F(-M_n - b_n z) - F(-M_n)|\, dK(z) + 1 - F(M_n) + F(-M_n). \tag{10}$$

Remark 3. Set $F_1 = F_n$ and $F_2 = F$. Then similarly we have

$$1 - F_n(M_n) + F_n(-M_n) \leqslant 2\|F_n - F\|_\infty + 1 - F(M_n) + F(-M_n). \tag{11}$$

Now we proceed to the main results. Let

$$c_n^*(t) = \mathbf{M}[c_n(t)] = \int\limits_{-\infty}^{\infty} e^{itx} dF_n^*(x).$$

The following theorem establishes the property of asymptotic uniform unbiasedness of the estimator $c_n(t)$.

Theorem 1. *Let F satisfy the Lipschitz condition*

$$|F(x) - F(y)| \leqslant L|x - y|.$$

Then as $n \to \infty$ $\|c_n - c\|_{T_{1n}}^{T_{2n}} \to 0$, where

$$T_n = \max\left(|T_{1n}|, |T_{2n}|\right) = o\left(b_n^{-1}\right).$$

Proof. According to Lemma 1 and relations (9), (10) we have

$$\|c_n - c\|_{T_{1n}}^{T_{2n}} \leqslant 2(1 + T_n M_n) \left\| \int_c^d \left(F(x - b_n z) - F(x)\right) dK(z) \right\|_\infty +$$

$$+ \int_c^d |F(x - b_n z) - F(M_n)|\, dK(z) + \int_c^d |F(-M_n - b_n z) - F(-M_n)|\, dK(z) +$$

$$+2[1 - F(M_n) + F(-M_n)].$$

Using the condition of the theorem and the estimate

$$\int_c^d |z| dK(z) \leqslant e(c, d), \tag{12}$$

we have

$$\|c_n^* - c\|_{T_{1n}}^{T_{2n}} \leqslant 2\left[(2 + T_n M_n) b_n \cdot L \cdot e(c, d) + 1 - F(M_n) + F(-M_n)\right].$$

Set $T_n b_n M_n = o(1)$, $n \to \infty$. Then we have the statement of the theorem. Theorem 1 is proved.

Remark 4. Let $T_n = b_n^{-\alpha}$, $0 < \alpha < 1$, and $M_n = b_n^{-\beta}$, $0 < \beta < 1 - \alpha$. Then obviously $T_n b_n M_n = b_n^{1-\alpha-\beta} = o(1)$ as $n \to \infty$.

In further results the sequence $\{b_n,\ n \geqslant 1\}$ may be random. Examples of such kernel d.f. are the estimates of "nearest neighbors." Let us prove uniformly strong consistency of c_n on finite interval $[T_1, T_2]$.

Theorem 2. *Let the condition* (11) *be satisfied and*

$$\frac{n b_n}{\log\ n} \xrightarrow{a.s.} \infty \quad as \quad n \to \infty. \tag{13}$$

Then

$$\|c_n - c\|_{T_1}^{T_2} \xrightarrow{a.s.} 0 \quad as \quad n \to \infty. \tag{14}$$

Proof. According to the triangle inequality

$$\|c_n - c\|_{T_1}^{T_2} \leqslant \left\|c_n - c_n^{\mathrm{e}}\right\|_{T_1}^{T_2} + \left\|c_n^{\mathrm{e}} - c\right\|_{T_1}^{T_2}.$$

From Theorem 1 in [5] we have

$$\left\|c_n^{\mathrm{e}} - c\right\|_{T_1}^{T_2} \xrightarrow{\text{a.s.}} 0, \text{ when} \quad n \to \infty. \tag{15}$$

Using condition of Theorem 1, relations (6)–(8) and (12), we have

$$\left\|c_n - c_n^{\mathrm{e}}\right\|_{T_1}^{T_2} \leqslant 2\Bigg\{(2 + T \cdot M_n)\left[\|Q_n(x)\|_\infty + L \cdot b_n \int\limits_c^d |z| dK(z)\right] +$$

$$+1 - F_n^{\mathrm{e}}(M_n) + F_n^{\mathrm{e}}(-M_n)\Bigg\} \overset{\text{a.s.}}{=} o\left(M_n b_n \left(\frac{\log n}{n b_n}\right)^{1/2}\right) + O(M_n b_n) + \varepsilon_n, \tag{16}$$

where we used inequality (12) and $\varepsilon_n = 1 - F_n^{\mathrm{e}}(M_n) + F_n^{\mathrm{e}}(-M_n)$. Let $M_n b_n \overset{\text{a.s.}}{=} o(1)$, $n \to \infty$. Then the statement of the theorem follows from (15) and (16). Theorem 2 is proved.

Note that in the proof of Theorem 2 M_n may be chosen as $M_n = b_n^{-\alpha}$, $0 < \alpha < 1$. In the next theorem we establish uniformly strong consistency of the estimate c_n on expanding interval $[T_{1n}, T_{2n}]$.

Theorem 3. *Let the conditions of Theorem* 2 *be satisfied, and the sequences of numbers* $\{T_{kn},\ n \geqslant 1,\ k = 1, 2\}$ *be such that when* $n \to \infty$

$$\begin{cases} T_n = \max\left(|T_{1n}|, |T_{2n}|\right) \overset{a.s.}{=} \left(\left(\dfrac{n}{\log\log n}\right)^{1/2}\right), \\ T_n \cdot b_n \xrightarrow{a.s.} 0. \end{cases} \tag{17}$$

Then

$$\|c_n - c\|_{T_1}^{T_2} \xrightarrow{a.s.} 0 \ \textit{when}\ n \to \infty.$$

Proof. As in the proof of Theorem 2, we use triangle inequality and with the first condition in (17) and Theorem 1 from [5] we have $\|c_n^{\mathrm{e}} - c\|_{T_{1n}}^{T_{2n}} \longrightarrow 0$, $n \to \infty$. According to the estimate (16),

$$\left\|c_n - c_n^{\mathrm{e}}\right\|_{T_{1n}}^{T_{2n}} \overset{\text{a.s.}}{=} o\left(T_n M_n b_n \left(\frac{n}{\log\log\ n}\right)^{1/2}\right) + O(T_n M_n b_n) + \varepsilon_n. \tag{18}$$

Now it remains to choose M_n to satisfy condition $T_n M_n b_n \xrightarrow{\text{a.s.}} 0$, $n \to \infty$. To do this we may set, for example, $T_n = M_n = n^{1/4}$, $b_n = n^{-\alpha}$, $\alpha > 1/2$. Theorem 3 is proved.

Theorems 2 and 3 may also be used to establish convergences of the form

$$a_n \left\|c_n - c_n^{\mathrm{e}}\right\|_{T_1}^{T_2} \xrightarrow{\text{a.s.}} 0 \text{ and } a_n \left\|c_n - c_n^{\mathrm{e}}\right\|_{T_{1n}}^{T_{2n}} \xrightarrow{\text{a.s.}} 0,\ \text{as } n \to \infty,$$

where the sequence of numbers $\{a_n,\ n \geqslant 1\}$ is such that $a_n \uparrow \infty$ when $n \to \infty$.

Theorem 4. *Let the conditions of Theorem* 2 *be satisfied. Then the following assertions hold*:

A. *If* $a_n b_n \xrightarrow{a.s.} 0$, *when* $n \to \infty$, *then*

$$a_n \cdot \left\|c_n - c_n^{e}\right\|_{T_1}^{T_2} \xrightarrow{a.s.} 0, \quad n \to \infty. \tag{19}$$

B. *If the first condition in* (17) *is satisfied and* $a_n b_n T_n \xrightarrow{a.s.} 0$, *when* $n \to \infty$, *then*

$$a_n \cdot \left\|c_n - c_n^{e}\right\|_{T_{1n}}^{T_{2n}} \xrightarrow{a.s.} 0, \quad n \to \infty. \tag{20}$$

Proof. Using (16), we obtain

$$a_n \cdot \|c_n - c_n\|_{T_1}^{T_2} \overset{\text{a.s.}}{=} o\left(a_n b_n M_n \left(\frac{\log n}{n b_n}\right)^{1/2}\right) + O(a_n b_n M_n) + a_n \varepsilon_n,\ n \to \infty.$$

Let us choose M_n to satisfy conditions $a_n b_n M_n \xrightarrow{\text{a.s.}} 0$ and $a_n \varepsilon_n \xrightarrow{\text{a.s.}} 0$, $n \to \infty$. Then (19) immediately follows. To prove (20) it suffices to use (18) and relations

$$M_n T_n a_n b_n \xrightarrow{\text{a.s.}} 0, \quad a_n \varepsilon_n \xrightarrow{\text{a.s.}} 0, \quad n \to \infty.$$

Theorem 4 is proved.

Remark 5. Let $a_n = n^{1/2}$, $b_n = cn^{-\alpha}$, $\alpha > 1/2$, $c > 0$. Then in the proof of (19) and (20) we may take $M_n \overset{\text{a.s.}}{=} o\left(n^{\alpha - 1/2}\right)$ and $M_n \overset{\text{a.s.}}{=} o\left(T_n^{-1} n^{\alpha-1/2}\right)$ respectively, such that $\varepsilon_n \overset{\text{a.s.}}{=} o\left(n^{-1/2}\right)$ when $n \to \infty$.

Remark 6. It is known [4, 5] that the complex-valued random process

$$\left\{Y_n^{\mathrm{e}}(t) = n^{1/2}\left(c_n(t) - c(t)\right), t \in \mathbb{R}\right\}$$

weakly converges in the Skorokhod space $D[T_1, T_2]$ $(-\infty < T_1 < T_2 < \infty)$ to the complex-valued Gaussian process $Y(t)$, where $\mathbf{M}Y(t) = 0$ and $\mathbf{M}Y(t)Y(s) = c(t+s) - c(t) \cdot c(s)$. Under conditions of Theorem 4 from (19) it follows that $Y_n^{\mathrm{e}}(t)$ and $Y_n(t) = n^{1/2}(c_n(t) - c(t))$ are identically distributed for $t \in [T_1, T_2]$ in the limit, when $n \to \infty$. This fact may be used to construct tests for goodness-of-fit hypothesis $H_0:\ F = F_0$ with the help of statistics

$$S_{1n}(Y_n) = \int_{T_1}^{T_2} |Y_n(t)|^2 dQ(t), \quad S_{2n}(Y_n) = \int_{T_1}^{T_2} (\operatorname{Re} Y_n(t))^2 dQ(t),$$

$$S_{3n}(Y_n) = \int_{T_1}^{T_2} (\operatorname{Im} Y_n(t))^2 dQ(t), \quad S_{4n}(Y_n) = \sup_{T_1 \leqslant t \leqslant T_2} |Y_n(t)|,$$

where $Q(t)$ is some d.f. with support in $[T_1, T_2]$.

Now we present the law of the iterated logarithm for the estimator c_n.

Theorem 5. *Under the conditions of Theorem 2 let*

$$F(-x) + 1 - F(x) = O\left(\frac{1}{x^\gamma}\right), \quad \gamma > 2, \quad x \to \infty.$$

Then as $n \to \infty$

$$\|c_n - c\|_{T_1}^{T_2} = O\left(\left(\frac{\log\log n}{n}\right)^{1/2-\alpha}\right),$$

where $\alpha = 1/(2(\gamma - 1))$.

Proof. Let $\mu_n = n/\log\log n$. Using estimates (2) and (11), we have

$$\|c_n - c\|_{T_1}^{T_2} \leqslant 2\left[(2 + TM_n)\|F_n - F\|_\infty + 1 - F(M_n) + F(-M_n)\right].$$

Let $M_n = \mu_n^\alpha$. Then according to (7)–(10) and law of the iterated logarithm for F_n^{e} we have

$$\mu_n^{1/2-\alpha}\|c_n - c\|_{T_1}^{T_2} \leqslant 2(2 + T\mu_n^\alpha)\mu_n^{-1/2}\mu_n^{1/2}\mu_n^{-\alpha} + O\left(\mu_n^{-\alpha\gamma} \cdot \mu_n^{1/2-\alpha}\right) \leqslant$$

$$\leqslant 4\mu_n^{-\alpha} + 2T + O\left(\mu_n^{-\frac{1}{\gamma-1}}\right) \to 2T, \quad n \to \infty.$$

Thus, Theorem 5 is proved.

Acknowledgements

This research is supported by RFBR, projects №05-01-00229, №04-01-00481.

REFERENCES

1. A. A. Abdushukurov and K. S. Sagidullaev, "Approximations of smoothed empirical processes," *Trans. Inst. Higher Educ.*, No. 1–2, 51–58 (2003).
2. A. A. Borovkov, *Mathematical Statistics*, Nauka, Moscow (1984).
3. D. D. Boos, "Rates of convergence of the distance between distribution function estimators," *Metrika*, **33**, 197–202 (1986).
4. A. Feuerverger and R. Mureika, "The empirical characteristic function and its applications," *Ann. Stat.*, **5**, No. 1, 88–97 (1977).
5. S. Csörgő, "Limit behavior of empirical characteristic function," *Ann. Probab.*, **9**, No. 1, 130–144 (1981).