UDC 519.21

E. Nadaraya, P. Babilua, G. Sokhadze (Iv. Javakhishvili Tbilisi State Univ., Georgia)

THE ESTIMATION OF A DISTRIBUTION FUNCTION BY AN INDIRECT SAMPLE

ОЦІНЮВАННЯ ФУНКЦІЇ РОЗПОДІЛУ З ВИКОРИСТАННЯМ НЕПРЯМОЇ ВИБІРКИ

The problem of estimation of a distribution function is considered when the observer has an access only to some indicator random values. Some basic asymptotic properties of the constructed estimates are studied. In this paper, the limit theorems are proved for continuous functionals related to the estimate of $\widehat{F}_n(x)$ in the space $C[a, 1-a]$, $0 < a < 1/2$.

Розглянуто задачу оцінювання функції розподілу у випадку, коли спостерігач має доступ лише до деяких індикаторних випадкових значень. Вивчено деякі базові асимптотичні властивості побудованих оцінок. У статті доведено граничні теореми для неперервних функціоналів щодо оцінки $\widehat{F}_n(x)$ у просторі $C[a, 1-a]$, $0 < a < 1/2$.

Let $X_1, X_2, \ldots, X_n$ be a sample of independent observations of a random non-negative value X with a distribution function $F(x)$. In problems of the theory of censored observations, sample values are pairs $Y_i = (X_i \wedge t_i)$ and $Z_i = I(Y_i = X_i)$, $i = \overline{1, n}$, where t_i are given numbers ($t_i \neq t_j$ for $i \neq j$) or random values independent of X_i, $i = \overline{1, n}$. Throughout the paper, $I(A)$ denotes the indicator of the set A.

Our present study deals with a somewhat different case: an observer has an access only to the values of random variables $\xi_i = I(X_i < t_i)$ with $t_i = c_F \dfrac{2i-1}{2n}$, $i = \overline{1, n}$, $c_F = \inf\{x \geq 0\colon F(x) = 1\} < \infty$.

The problem consists in estimating the distribution function $F(x)$ by means of a sample $\xi_1, \xi_2, \ldots, \xi_n$. Such a problem arises for example from a region of corrosion investigations, see [1] where an experiment related to corrosion is described.

As an estimate for $F(x)$ we consider an expression of the form

$$\widehat{F}_n(x) = \begin{cases} 0, & x \leq 0, \\ F_{1n}(x) \cdot F_{2n}^{-1}(x), & 0 < x < c_F, \\ 1, & x \geq c_F, \end{cases} \tag{1}$$

$$F_{1n}(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - t_j}{h}\right) \xi_j,$$

$$F_{2n}(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - t_j}{h}\right),$$

where $K(x)$ is a probability density (kernel), $K(x) = K(-x)$, $x \in (-\infty, \infty)$, $\{h = h(n)\}$ is a sequence of positive numbers converging to zero.

1. In this subsection we give the conditions of asymptotic unbiasedness and consistency and the theorems on a limiting distribution $\widehat{F}_n(x)$.

© E. NADARAYA, P. BABILUA, G. SOKHADZE, 2010

Lemma 1. *Assume that*

1^0. $K(x)$ *is a function of bounded variation. If* $nh \to \infty$, *then*

$$\frac{1}{nh}\sum_{j=1}^{n} K^{m_1-1}\left(\frac{x-t_j}{h}\right) F^{m_2-1}(t_j) =$$

$$= \frac{1}{c_F h}\int_0^{c_F} K^{m_1-1}\left(\frac{x-u}{h}\right) F^{m_2-1}(u)\,du + O\left(\frac{1}{nh}\right), \tag{2}$$

uniformly with respect to $x \in [0, c_F]$; m_1, m_2 *are natural numbers.*

Proof. Let $P(x)$ be a uniform distribution function on $[0, c_F]$, and $P_n(x)$ be an empirical distribution function of "the sample" $t_1, t_2, \ldots, t_n$, i.e., $P_n(x) = n^{-1}\sum_{j=1}^{n} I(t_j < x)$. It is obvious that

$$\sup_{0\le x\le c_F} |P_n(x) - P(x)| = \sup_{0\le x\le c_F}\left|\frac{1}{n}\left[n\frac{x}{c_F} + \frac{1}{2}\right] - \frac{x}{c_F}\right| \le \frac{1}{2n}\,. \tag{3}$$

We have

$$\frac{1}{nh}\sum_{i=1}^{n} K^{m_1-1}\left(\frac{x-t_i}{h}\right) F^{m_2-1}(t_i) -$$

$$-\frac{1}{c_F h}\int_0^{c_F} K^{m_1-1}\left(\frac{x-u}{h}\right) F^{m_2-1}(u)\,du =$$

$$= \frac{1}{h}\int_0^{c_F} K^{m_1-1}\left(\frac{x-u}{h}\right) F^{m_2-1}(u)\,d(P_n(u) - P(u)). \tag{4}$$

Applying the integration by parts formula to the integral in the right-hand part of (4) and taking (3) into account, we obtain (2).

Lemma 1 is proved.

Below it is assumed without loss of generality that the interval $[0, c_F] = [0, 1]$.

Theorem 1. *Let* $F(x)$ *be continuous and the conditions of the Lemma* 1 *be fulfilled. Then the estimate* (1) *is asymptotically unbiased and consistent at all points* $x \in [0, 1]$. *Moreover,* $\widehat{F}_n(x)$ *has an asymptotically normal distribution, i.e.,*

$$\sqrt{nh}\left(\widehat{F}_n(x) - E\widehat{F}_n(x)\right)\sigma^{-1}(x) \xrightarrow{d} N(0,1),$$

$$\sigma^2(x) = F(x)(1 - F(x))\int K^2(u)\,du,$$

where d *denotes convergence in distribution, and* $N(0,1)$ *a random value having a normal distribution with mean* 0 *and variance* 1.

Proof. By Lemma 1 we have

$$EF_{1n}(x) = \int\limits_{\frac{x-1}{h}}^{\frac{x}{h}} K(t)F(x+ht)\,dt + O\left(\frac{1}{nh}\right),$$

$$F_{2n}(x) = \frac{1}{h}\int\limits_{0}^{1} K\left(\frac{x-u}{h}\right)du + O\left(\frac{1}{nh}\right), \tag{5}$$

and for $n \to \infty$

$$\frac{1}{h}\int\limits_{0}^{1} K\left(\frac{x-u}{h}\right)du \longrightarrow F_2(x) = \begin{cases} 1, & x \in (0,1), \\ \frac{1}{2}, & x = 0, \quad x = 1, \end{cases}$$

$$\int\limits_{\frac{x-1}{h}}^{\frac{x}{h}} K(t)F(x+th)\,dt \longrightarrow F(x)F_2(x).$$

Hence it follows that $E\widehat{F}_n(x) \to F(x)$, $x \in [0,1]$ as $n \to \infty$.

Analogously, it is not difficult to show that

$$\operatorname{Var}\widehat{F}_n(x) =$$

$$= \left[\frac{1}{nh^2}\int\limits_{0}^{1} K^2\left(\frac{x-u}{h}\right)F(u)(1-F(u))\,du + O\left(\frac{1}{(nh)^2}\right)\right]F_{2n}^{-2}(x).$$

Hence we readily derive

$$nh\operatorname{Var}\widehat{F}_n(x) \sim \sigma^2(x) = F(x)(1-F(x))\int K^2(u)\,du \tag{6}$$

for $x \in [0,1]$.

Thus $\widehat{F}_n(x)$ is a consistent estimate for $F(x)$, $x \in [0,1]$, and therefore

$$P\left\{\widehat{F}_n(x_1) \le \widehat{F}_n(x_2)\right\} \longrightarrow 1 \quad \text{as} \quad n \to \infty, \quad x_1 < x_2, \quad x_1,\, x_2 \in [0,1].$$

Let us now establish that $\widehat{F}_n(x)$ has an asymptotically normal distribution. Since, by virtue of (5), $F_{2n}(x) \to F_2(x)$, it remains for us to verify the condition of the Liapunov central limit theorem for $F_{1n}(x)$.

Let us denote

$$\eta_i = \eta_i(x) = (nh)^{-1}K\left(\frac{x-t_i}{h}\right)\xi_i$$

and show that

$$L_n = \sum_{j=1}^{n} E|\eta_j - E\eta_j|^{2+\delta}(\operatorname{Var}F_{1n}(x))^{-1-\frac{\delta}{2}} \longrightarrow 0, \quad \delta > 0. \tag{7}$$

We have

$$\sum_{j=1}^{n} E|\eta_j - E\eta_j|^{2+\delta} \leq 2M^{1+\delta}(nh)^{-(2+\delta)} \sum_{j=1}^{n} K\left(\frac{x-t_j}{h}\right) F(t_j),$$

$$M = \max_{x \in R} K(x).$$

Hence, taking (2) into account, we find

$$\sum_{i=1}^{n} E|\eta_i - E\eta_i|^{2+\delta} \leq c_1(nh)^{-(1+\delta)}. \tag{8}$$

Using the relation (6) and the inequality (8), we establish that $L_n = O((nh)^{-\frac{\delta}{2}})$, i.e., (7) holds.

Theorem 1 is proved.

2. Uniform consistency. In this subsection we define the conditions, under which the estimate $\widehat{F}_n(x)$ uniformly converges in probability (a. s.) to true $F(x)$.

Let us introduce the Fourier transform of the function $K(x)$

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} K(x)\, dx$$

and assume that

2^0. $\varphi(t)$ is absolutely integrable. Following E. Parzen [2], $F_{1n}(x)$ can be represented as

$$F_{1n}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iu\frac{x}{h}} \varphi(u) \frac{1}{nh} \sum_{j=1}^{n} \xi_j e^{iu\frac{t_j}{h}}\, du.$$

Thus

$$F_{1n}(x) - EF_{1n}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iu\frac{x}{h}} \varphi(u) \frac{1}{nh} \sum_{j=1}^{n} (\xi_j - F(t_j)) e^{iu\frac{t_j}{h}}\, du.$$

Denote

$$d_n = \sup_{x \in \Omega_n} |\widehat{F}_n(x) - E\widehat{F}_n(x)|, \quad \Omega_n = [h^\alpha, 1 - h^\alpha], \quad 0 < \alpha < 1.$$

Theorem 2. *Let $K(x)$ satisfy conditions 1^0 and 2^0.*

(a) *Let $F(x)$ be continuous and $n^{\frac{1}{2}} h_n \to \infty$, then*

$$D_n = \sup_{x \in \Omega_n} |\widehat{F}_n(x) - F(x)| \xrightarrow{P} 0.$$

(b) *If $\sum_{n=1}^{\infty} n^{-\frac{p}{2}} h^{-p} < \infty$, $p > 2$, then $D_n \to 0$ a. s.*

Proof. We have

$$\sup_{x \in \Omega_n} \left(1 - \frac{1}{h}\int_0^1 K\left(\frac{x-u}{h}\right) du\right) \leq \int_{-\infty}^{-h^{\alpha-1}} K(u)\, du + \int_{h^{\alpha-1}}^{\infty} K(u)\, du \longrightarrow 0. \tag{9}$$

This and (5) imply that

$$\sup_{x\in\Omega_n} |F_{2n}(x) - 1| \longrightarrow 0, \tag{10}$$

i.e., due to the uniform convergence for any $\varepsilon_0 > 0$, $0 < \varepsilon_0 < 1$, and sufficiently large $n \geq n_0$, we have $F_{2n}(x) \geq 1 - \varepsilon_0$ uniformly with respect to $x \in \Omega_n$. Therefore,

$$d_n \leq (1-\varepsilon_0)^{-1} \sup_{x\in\Omega_n} |F_{1n}(x) - EF_{1n}(x)| \leq$$

$$\leq (1-\varepsilon_0)^{-1} \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} |\varphi(u)| \frac{1}{nh} \left| \sum_{j=1}^{n} \overline{\eta}_j e^{iu\,\frac{t_j}{h}} \right| du, \quad \overline{\eta}_j = \xi_j - F(t_j).$$

Hence, by Hölder's inequality, we obtain

$$d_n^p \leq (1-\varepsilon_0)^{-p} \frac{1}{(2\pi)^p} \int\limits_{-\infty}^{\infty} |\varphi(u)| \left| \sum_{j=1}^{n} \overline{\eta}_j e^{iu\,\frac{t_j}{h}} \right|^p du \left(\int\limits_{-\infty}^{\infty} |\varphi(u)|\, du \right)^{\frac{p}{q}},$$

$$\frac{1}{p} + \frac{1}{q} = 1, \quad p > 2.$$

Thus

$$Ed_n^p \leq c(\varepsilon, p, \varphi)\, \frac{1}{(nh)^p} \int\limits_{-\infty}^{\infty} |\varphi(u)| E \left| \sum_{j,k} \cos\left(\left(\frac{t_j - t_k}{h} \right) u \right) \overline{\eta}_j \overline{\eta}_k \right|^{\frac{p}{2}} du, \tag{11}$$

where

$$c(\varepsilon, p, \varphi) = (1-\varepsilon_0)^{-p}\, \frac{1}{(2\pi)^p} \left(\int\limits_{-\infty}^{\infty} |\varphi(u)|\, du \right)^{\frac{p}{q}}.$$

Denote

$$A(u) = \sum_{j,k} \cos\left(\left(\frac{t_j - t_k}{h} \right) u \right) \overline{\eta}_j \overline{\eta}_k.$$

Then by (11) we write

$$Ed_n^p \leq 2^{\frac{p}{2}-1} c(\varepsilon_0, p, \varphi)\, \frac{1}{(nh)^p} \times$$

$$\times \left[\int\limits_{-\infty}^{\infty} |\varphi(u)|\, |EA(u)|^{\frac{p}{2}}\, du + \int\limits_{-\infty}^{\infty} |\varphi(u)|\, E|A(u) - EA(u)|^{\frac{p}{2}}\, du \right]. \tag{12}$$

Using Whittle's inequality [3] for moments of quadratic form, we obtain

$$E|A(u) - EA(u)|^{\frac{p}{2}} \leq$$

$$\leq 2^{\frac{3}{2}\,p} c\left(\frac{p}{2}\right) [c(p)]^{\frac{1}{2}} \left(\sum_{i,j} \cos^2\left(\left(\frac{t_j - t_k}{h} \right) u \right) \gamma_j^2(p) \gamma_k^2(p) \right)^{\frac{p}{4}},$$

where

$$\gamma_k(p) = (E|\overline{\eta}_k|^p)^{\frac{1}{p}} \le 1, \quad c(p) = \frac{2^{\frac{p}{2}}}{\sqrt{\pi}}\,\Gamma\left(\frac{p+1}{2}\right).$$

Hence it follows that

$$E|A(u) - EA(u)|^{\frac{p}{2}} = O(n^{\frac{p}{2}}) \tag{13}$$

uniformly with respect to $u \in (-\infty, \infty)$. It is also clear that

$$|EA(u)|^{\frac{p}{2}} = O(n^{\frac{p}{2}}) \tag{14}$$

uniformly with respect to $u \in (-\infty, \infty)$.

Having combined the relations (12), (13) and (14), we obtain

$$Ed_n^p = O\left(\frac{1}{(\sqrt{n}\,h)^p}\right), \quad p > 2.$$

Therefore,

$$P\left\{\sup_{x\in\Omega_n}\left|\widehat{F}_n(x) - E\widehat{F}_n(x)\right| \ge \varepsilon\right\} \le \frac{c_3}{\varepsilon^p(\sqrt{n}\,h)^p}\,. \tag{15}$$

Furthermore, we have

$$\sup_{x\in\Omega_n}\left|E\widehat{F}_n(x) - F(x)\right| \le$$

$$\le \frac{1}{1-\varepsilon_0}\left(\sup_{x\in\Omega_n}|EF_{1n}(x) - F(x)| + \sup_{x\in\Omega_n}|1 - F_{2n}(x)|\right). \tag{16}$$

By virtue of (10), the second summand in the right-hand part of (16) tends to 0, whereas the first summand is estimated as follows:

$$\sup_{x\in\Omega_n}|EF_{1n}(x) - F(x)| \le S_{1n} + S_{2n} + O\left(\frac{1}{nh}\right), \tag{17}$$

$$S_{1n} = \sup_{0\le x\le 1}\left|\frac{1}{h}\int_0^1 (F(y) - F(x))K\left(\frac{x-y}{h}\right)dy\right|,$$

$$S_{2n} = \sup_{x\in\Omega_n}\left(1 - \frac{1}{h}\int_0^1 K\left(\frac{x-y}{h}\right)dy\right),$$

and, by virtue of (9),

$$S_{2n} \longrightarrow 0 \tag{18}$$

as $n \to \infty$.

Let us now consider S_{1n}. Note that

$$S_{1n} \le \sup_{0\le x\le 1}\left|\int_0^1 |F(y) - F(x)|\,\frac{1}{h}\,K\left(\frac{x-y}{h}\right)\right| dy =$$

$$= \sup_{0\le x\le 1} \int\limits_{x-1}^{x} |F(x-u)-F(x)| \, \frac{1}{h} K\left(\frac{u}{h}\right) du \le$$

$$\le \sup_{0\le x\le 1} \int\limits_{-\infty}^{\infty} |F(x-u)-F(x)| \, \frac{1}{h} K\left(\frac{u}{h}\right) du. \tag{19}$$

Assume that $\delta > 0$ and divide the integration domain in (19) into two domains $|u| \le \delta$ and $|u| > \delta$. Then

$$S_{1n} \le \sup_{0\le x\le 1} \int\limits_{|u|\le\delta} |F(x-u)-F(x)| \, \frac{1}{h} K\left(\frac{u}{h}\right) du+$$

$$+ \sup_{0\le x\le 1} \int\limits_{|u|>\delta} |F(x-u)-F(x)| \, \frac{1}{h} K\left(\frac{u}{h}\right) du \le$$

$$\le \sup_{x\in R} \sup_{|u|\le\delta} |F(x-u)-F(x)| + 2 \int\limits_{|u|\ge\frac{\delta}{h}} K(u)\, du. \tag{20}$$

By a choice of $\delta > 0$ the first summand in the right-hand part of (20) can be made arbitrarily small. Choosing $\delta > 0$ and letting $n \to \infty$, we find that the second summand tends to zero. Therefore,

$$\lim_{n\to\infty} S_{1n} = 0. \tag{21}$$

Finally, from the relations (15) – (18) and (21) the proof of the theorem follows.

***Remark* 1.** 1. If $K(x) = 0$, $|x| \ge 1$ and $\alpha = 1$, i. e., $\Omega_n = [h, 1-h]$, then $S_{2n} = 0$.

2. In the conditions of Theorem 2

$$\sup_{x\in[a,b]} |\widehat{F}_n(x) - F(x)| \longrightarrow 0$$

in probability (a. s.) for any fixed interval $[a, b] \subset [0, 1]$ since there may exist n_0 such that $[a, b] \subset \Omega_n$, $n \ge n_0$.

Assume that $h = n^{-\gamma}$, $\gamma > 0$. The conditions of Theorem 2 are fulfilled: $n^{\frac{1}{2}} h_n \to \to \infty$ if $0 < \gamma < \dfrac{1}{2}$, and

$$\sum_{n=1}^{\infty} n^{-\frac{p}{2}} h_n^{-p} < \infty \quad \text{if} \quad 0 < \gamma < \frac{p-2}{2p}, \quad p > 2.$$

3. Estimation of moments. In considering the problem, there naturally arises a question of estimation of the integral functionals of $F(x)$, for example, moments μ_m, $m \ge 1$:

$$\mu_m = m \int\limits_0^1 t^{m-1}(1 - F(t))\, dt.$$

As estimates for μ_m we consider the statistics

$$\widehat{\mu}_{nm} = 1 - \frac{m}{n} \sum_{j=1}^{n} \xi_j \, \frac{1}{h} \int_{h}^{1-h} t^{m-1} K\left(\frac{t - t_j}{h}\right) F_{2n}^{-1}(t)\, dt.$$

Theorem 3. *Let $K(x)$ satisfy condition 1^0 and, in addition to this, $K(x) = 0$ outside the interval $[-1, 1]$. If $nh \to \infty$ as $n \to \infty$, then $\widehat{\mu}_{nk}$ is an asymptotically unbiased, consistent estimate for μ_m and moreover*

$$\frac{\sqrt{n}\,(\widehat{\mu}_{nm} - E\widehat{\mu}_{nm})}{\sigma} \xrightarrow{\ d\ } N(0,1), \quad \sigma^2 = m^2 \int_0^1 t^{2m-2} F(t)(1 - F(t))\, dt.$$

Proof. Since $K(x)$ has $[-1, 1]$ as a support, we establish from (5) that $F_{2n}(x) = 1 + O\left(\dfrac{1}{nh}\right)$ uniformly with respect to $x \in [h, 1-h]$.

Hence, by Lemma 1 we have

$$E\widehat{\mu}_{nm} = 1 - \frac{m}{n} \sum_{j=1}^{n} F(t_j) \, \frac{1}{h} \int_{h}^{1-h} t^{m-1} K\left(\frac{t - t_j}{h}\right) F_{2n}^{-1}(t)\, dt =$$

$$= 1 - m \int_{h}^{1-h} \left[\frac{1}{h} \int_0^1 K\left(\frac{t-u}{h}\right) F(u)\, du \right] t^{m-1}\, dt + O\left(\frac{1}{nh}\right) =$$

$$= 1 - m \int_{h}^{1-h} \left(\int_{-1}^{1} K(v) F(t + vh)\, dv \right) t^{m-1}\, dt + O\left(\frac{1}{nh}\right) =$$

$$= 1 - m \int_0^1 t^{m-1} \left[\int_{-1}^{1} K(v) F(t + vh)\, dv \right] dt + O(h) + O\left(\frac{1}{nh}\right). \qquad (22)$$

By the Lebesgue theorem on majorized convergence, from (22) we establish that

$$E\widehat{\mu}_{nm} \longrightarrow 1 - m \int_0^1 F(t) t^{m-1}\, dt =$$

$$= m \int_0^1 t^{m-1} (1 - F(t))\, dt = \mu_m, \quad m \geq 1. \qquad (23)$$

Therefore, $\widehat{\mu}_{nm}$ is an asymptotically unbiased estimate for μ_m.

Further, analogously to (22), it can be shown that

$$\operatorname{Var} \widehat{\mu}_{nm} = \frac{m^2}{n} \int_0^1 F(t)(1 - F(t)) t^{2m-2} \left[\mathcal{K}\left(\frac{1-t}{h} - 1\right) - \mathcal{K}\left(1 - \frac{t}{h}\right) \right]^2 dt +$$

$$+ O\left(\frac{h}{n}\right) + O\left(\frac{1}{(nh)^2}\right),$$

where

$$\mathcal{K}(v) = \int_{-\infty}^{v} K(u)\,du.$$

By the same Lebesgue theorem we see that

$$n\,\mathrm{Var}\,\widehat{\mu}_{nm} \sim \sigma^2 = m^2 \int_0^1 t^{2m-2} F(t)(1-F(t))\,dt. \tag{24}$$

Therefore (23) and (24) imply that $\widehat{\mu}_{nm} \xrightarrow{P} \mu_m$.

To complete the proof of the theorem it remains to show that the statistics $\sqrt{n}\,(\widehat{\mu}_{nm} - E\widehat{\mu}_{nm})$ have an asymptotically normal distribution with mean 0 and dispersion σ^2. For this it suffices to show that the Liapunov fraction $L_n \to 0$. Indeed,

$$L_n = n^{-(2+\delta)} m^{2+\delta} \sum_{j=1}^{n} E|\xi_j - F(t_j)|^{2+\delta} \times$$

$$\times \left| \frac{1}{h} \int_h^{1-h} t^{m-1} K\left(\frac{t-t_j}{h}\right) F_{2n}^{-1}(t)\,dt \right|^{2+\delta} (\mathrm{Var}\,\widehat{\mu}_{nm})^{-(1+\frac{\delta}{2})} \le$$

$$\le c_6 n^{-(2+\delta)} \sum_{j=1}^{n} E|\xi_j - F(t_j)|^{2+\delta} (\mathrm{Var}\,\widehat{\mu}_{nm})^{-(1+\frac{\delta}{2})} \le$$

$$\le c_7 n^{-1-\delta} (\mathrm{Var}\,\widehat{\mu}_{nm})^{-1-\frac{\delta}{2}} = O(n^{-\frac{\delta}{2}}).$$

Theorem 3 is proved.

4. Limit theorems of functionals related to the estimate $\widehat{F}_n(x)$. In this subsection the kernel $K(x) \ge 0$ is chosen so that it would be a function of finite variation and satisfy the conditions

$$K(-u) = K(u), \quad \int K(u)\,du = 1,$$

$$K(u) = 0 \ \text{ for } \ |u| \ge 1.$$

Theorem 4. *Let $g(x) \ge 0$, $x \in [a, 1-a]$, $0 < a < \frac{1}{2}$, be a measurable and bounded function.*

(a) *If $F(a) > 0$ and $nh^2 \to \infty$ as $n \to \infty$, then*

$$\overline{T}_n = \sqrt{n} \int_a^{1-a} g_1(x) \left[\widehat{F}_n(x) - E\widehat{F}_n(x)\right] dx \xrightarrow{d} N(0, \sigma^2), \tag{25}$$

where

$$g_1(x) = g(x)\psi(F(x)), \quad \psi(t) = \frac{1}{\sqrt{t(1-t)}}.$$

(b) *If* $F(a) > 0$, $nh^2 \to \infty$, $nh^4 \to 0$ *as* $n \to \infty$ *and* $F(x)$ *has bounded derivatives up to second order, then as* $n \to \infty$

$$T_n = \sqrt{n} \int_a^{1-a} g_1(x) \left[\widehat{F}_n(x) - F(x)\right] dx \xrightarrow{d} N(0, \sigma^2),$$

$$\sigma^2 = \int_a^{1-a} g^2(u)\, du.$$

***Remark* 2.** We have introduced $a > 0$ in (25) in order to avoid the boundary effect of the estimate $\widehat{F}_n(x)$ since near the interval boundary the estimate $\widehat{F}_n(x)$ being a kernel type estimate behaves worse in the sense of order of bias tendency to zero than on any inner interval $[a, 1-a] \subset [0, 1]$, $0 < a < \dfrac{1}{2}$.

***Proof of Theorem* 4.** We have

$$\overline{T}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\xi_j - F(t_j)) \frac{1}{h} \int_a^{1-a} K\left(\frac{u - t_j}{h}\right) g_{2n}(u)\, du,$$

where

$$g_{2n}(u) = g_1(u) F_{2n}^{-1}(u).$$

Hence

$$\sigma_n^2 = \operatorname{Var} \overline{T}_n =$$

$$= \frac{1}{n} \sum_{j=1}^{n} \psi^{-2}(F(t_j)) \left(\frac{1}{h} \int_a^{1-a} K\left(\frac{u - t_j}{h}\right) g_{2n}(u)\, du\right)^2. \qquad (26)$$

Since $K(u)$ has $[-1, 1]$ as a support and $0 < a \le u \le 1 - a$, it can be easily verified that

$$F_{2n}(u) = 1 + O\left(\frac{1}{nh}\right) \quad \text{and} \quad g_{2n}(u) = g_1(u) + O\left(\frac{1}{nh}\right)$$

uniformly on $u \in [a, 1-a]$. Therefore, from (26) we have

$$\sigma_n^2 = \frac{1}{n} \sum_{j=1}^{n} \psi^{-2}(F(t_j)) \left(\frac{1}{h} \int_a^{1-a} K\left(\frac{u - t_j}{h}\right) g_1(u)\, du\right)^2 + O\left(\frac{1}{nh}\right).$$

By virtue of Lemma 1, we can easily show that

$$\frac{1}{n} \sum_{j=1}^{n} \psi^{-2}(F(t_j)) \left(\frac{1}{h} \int_a^{1-a} K\left(\frac{u - t_j}{h}\right) g_1(u)\, du\right)^2 =$$

$$= \int_0^1 \psi^{-2}(F(t))\, dt \left(\frac{1}{h} \int_a^{1-a} K\left(\frac{u - t}{h}\right) g_1(u)\, du\right)^2 + O\left(\frac{1}{nh^2}\right).$$

Therefore,

$$\sigma_n^2 = \int_a^{1-a} \psi^{-2}(F(t))\,dt \left(\frac{1}{h}\int_a^{1-a} K\left(\frac{u-t}{h}\right) g_1(u)\,du\right)^2 +$$

$$+\varepsilon_n^{(1)} + \varepsilon_n^{(2)} + O\left(\frac{1}{nh^2}\right), \tag{27}$$

$$\varepsilon_n^{(1)} = \int_0^{a} \psi^{-2}(F(t))\,dt \left(\frac{1}{h}\int_a^{1-a} K\left(\frac{u-t}{h}\right) g_1(u)\,du\right)^2,$$

$$\varepsilon_n^{(2)} = \int_{1-a}^{1} \psi^{-2}(F(t))\,dt \left(\frac{1}{h}\int_a^{1-a} K\left(\frac{u-t}{h}\right) g_1(u)\,du\right)^2.$$

Since by $F(u)(1-F(u)) \le \dfrac{1}{4}$, $g(u) \le c_8$ and

$$\psi(F(u)) \le \frac{1}{F(a)(1-F(1-a))}, \quad a \le u \le 1-a,$$

it follows that $g_1(u) \le c_9$, we have

$$\varepsilon_n^{(1)} \le c_{10}\int_0^a dt \left(\int_{\frac{a-t}{h}}^{\frac{1-a-t}{h}} K(u)\,du\right)^2, \tag{28}$$

where $a-t \ge 0$ and $1-a-t \ge 0$. The first inequality is obvious, whereas the second one follows from the inequalities $0 \le t \le a$ and $0 < a < \dfrac{1}{2}$.

Therefore,

$$\lim_{n\to\infty} \int_{\frac{a-t}{h}}^{\frac{1-a-t}{h}} K(u)\,du = \begin{cases} 0, & 0 \le t < a, \\ \dfrac{1}{2}, & t = a. \end{cases}$$

By the Lebesgue theorem on bounded convergence, from the latter expression and (28) we obtain

$$\varepsilon_n^{(1)} \to 0 \ \text{ as } \ n \to \infty. \tag{29}$$

Analogously,

$$\varepsilon_n^{(2)} \to 0 \ \text{ as } \ n \to \infty. \tag{30}$$

Now let us establish that

$$\int_a^{1-a} \psi^{-2}(F(t))\,dt \left(\frac{1}{h}\int_a^{1-a} K\left(\frac{u-t}{h}\right) g_1(u)\,du\right)^2 \longrightarrow \sigma^2 = \int_a^{1-a} g^2(u)\,du$$

as $n \to \infty$.

We have

$$\left|\int\limits_a^{1-a}\psi^{-2}(F(t))\,dt\left(\frac{1}{h}\int\limits_a^{1-a}g_1(u)K\left(\frac{u-t}{h}\right)\,du\right)^2-\int\limits_a^{1-a}\psi^{-2}(F(t))g_1^2(t)\,dt\right|\leq$$

$$\leq c_{11}\int\limits_a^{1-a}\psi^{-2}(F(t))\,dt\left|\frac{1}{h}\int\limits_a^{1-a}g_1(u)K\left(\frac{u-t}{h}\right)\,du-g_1(t)\right|\leq$$

$$\leq c_{12}\int\limits_a^{1-a}dt\left|\frac{1}{h}\int\limits_a^{1-a}g_1(u)K\left(\frac{u-t}{h}\right)\,du-g_1(t)\int\limits_a^{1-a}\frac{1}{h}\,K\left(\frac{u-t}{h}\right)\,du\right|+$$

$$+c_{13}\int\limits_a^{1-a}\left|\int\limits_a^{1-a}\frac{1}{h}\,K\left(\frac{u-t}{h}\right)\,du-1\right|dt=A_{1n}+A_{2n}. \qquad (31)$$

Since

$$\int\limits_a^{1-a}\frac{1}{h}\,K\left(\frac{u-t}{h}\right)\,du\longrightarrow 1$$

for all $t\in(a,1-a)$, we have

$$A_{2n}\to 0 \ \text{ as } \ n\to\infty. \qquad (32)$$

Further, we continue the function $g_1(u)$ so that that outside $[a,1-a]$ it has zero values and denote the continued function by $\overline{g}_1(u)$. Then

$$A_{1n}\leq c_{14}\left|\int\limits_0^1\left(\int\limits_{-\infty}^{\infty}|\overline{g}_1(x+y)-\overline{g}_1(y)|\;dy\right)\,\frac{1}{h}\,K\left(\frac{x}{h}\right)\right|\,dx\leq$$

$$\leq c_{15}\int\limits_{-1}^{1}\left(\int\limits_{-\infty}^{\infty}|\overline{g}_1(y+uh)-\overline{g}_1(y)|\;dy\right)K(u)\,du=$$

$$=c_{15}\int\limits_{-1}^{1}\omega(uh)K(u)\,du\longrightarrow 0 \ \text{ as } \ n\to\infty, \qquad (33)$$

where

$$\omega(y)=\int\limits_{-\infty}^{\infty}|\overline{g}_1(y+x)-\overline{g}_1(x)|\;dx.$$

The (33) holds by virtue of the Lebesgue theorem on majorized convergence and the fact that $\omega(uh)\leq 2\|\overline{g}_1\|_{L_1(-\infty,\infty)}$ and $\omega(uh)\to 0$ as $n\to\infty$. Thereby, taking

(27) – (33) into account, we have proved that

$$\sigma_n^2 \longrightarrow \sigma^2 = \int\limits_a^{1-a} g^2(u)\,du. \tag{34}$$

Now let us verify the fulfillment of the conditions of the central limit theorems for the sums

$$\overline{T}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_{jn}(\xi_j - F(t_j)),$$

$$a_{jn} = \int\limits_a^{1-a} \frac{1}{h} K\left(\frac{x - t_j}{h}\right) g_{2n}(x)\,dx.$$

We have

$$L_n = \frac{n^{-(1+\frac{\delta}{2})} \sum\limits_{j=1}^{n} a_{jn}^{2+\delta} E|\xi_j - F(t_j)|^{2+\delta}}{(\sqrt{\operatorname{Var} \overline{T}_n})^{2+\delta}} = O(n^{-\frac{\delta}{2}}),$$

since $a_{jn} \le c_{16}$, $E|\xi_j - F(t_j)|^{2+\delta} \le 1$ for all $1 \le j \le n$ and $\operatorname{Var} \overline{T}_n \to \sigma^2$.

Finally, the statement b) of the theorem follows from a) if we take into account that

$$\sqrt{n} \int\limits_a^{1-a} g_1(x) \left[E\widehat{F}_n(x) - F(x)\right] dx =$$

$$= \sqrt{n} \int\limits_a^{1-a} g_1(x) \left[\int\limits_{-1}^{1} \left[K(u)(F(x-uh) - F(x))\right] du\right] dx =$$

$$= O(\sqrt{n}\,h^2) + O\left(\frac{1}{\sqrt{n}\,h}\right). \tag{35}$$

Theorem 4 is proved.

Lemma 2. 1. *In the conditions of the item* (a) *of Theorem* 4,

$$E|\overline{T}_n|^s \le c_{17} \left(\int\limits_a^{1-a} g(u)\,du\right)^{\frac{s}{2}}, \quad s > 2. \tag{36}$$

2. *In the conditions of the item* (b) *of Theorem* 4,

$$E|T_n|^s \le c_{18} \left(\int\limits_a^{1-a} g(u)\,du\right)^{\frac{s}{2}}, \quad s > 2. \tag{37}$$

Proof. $\overline{T}_n$ is the linear form of $\eta_j = \xi_j - F(t_j)$, $E\eta_j = 0$, $1 \le j \le n$. Hence to prove (36) we use Whittle's inequality [3].

It is obvious that $E|\eta_j|^s \le 1$, $j = \overline{1,n}$. Therefore by Whittle's inequality

$$E|\overline{T}_n|^s \le c(s)2^s \left[\frac{1}{nh^2} \sum_{j=1}^{n} \left(\int_a^{1-a} K\left(\frac{u-t_j}{h}\right) g_{2n}(u)\, du \right)^2 \right]^{\frac{s}{2}},$$

where $g_{2n}(u) = g_1(u) F_{2n}^{-1}(u)$.

This, by virtue of Lemma 1, yields

$$E|\overline{T}_n|^s \le c(s)2^s \left[\int_0^1 \left(\frac{1}{h} \int_a^{1-a} K\left(\frac{u-t}{h}\right) g_{2n}(u)\, du \right)^2 dt + \right.$$

$$\left. + O\left(\frac{1}{nh^2}\right) \left(\int_a^{1-a} g_{2n}(u)\, du \right)^2 \right]^{\frac{s}{2}}. \tag{38}$$

Further, since

$$g_{2n}(u) \le g(u) \left[\frac{1}{F(a)(1-F(1-a))} \right] \left[1 + O\left(\frac{1}{nh}\right) \right] \le$$

$$\le c_{19} g(u), \quad a \le u \le 1-a,$$

from (38) it follows that

$$E|\overline{T}_n|^s \le c_{20} \left[\sup_{0 \le t \le 1} \left(\frac{1}{h} \int_a^{1-a} K\left(\frac{u-t}{h}\right) g_{2n}(u)\, du \right) \times \right.$$

$$\left. \times \int_0^1 dt \int_a^{1-a} \frac{1}{h} K\left(\frac{u-t}{h}\right) g_{2n}(u)\, du \right]^{\frac{s}{2}} + O\left(\frac{1}{nh^2}\right)^{\frac{s}{2}} \left(\int_a^{1-a} g_{2n}(u)\, du \right)^{\frac{s}{2}} \le$$

$$\le c_{21} \left(\int_a^{1-a} g(u)\, du \right)^{\frac{s}{2}} [1 + o(1)] \le c_{22} \left(\int_a^{1-a} g(u)\, du \right)^{\frac{s}{2}}, \quad s > 2.$$

Next we obtain

$$E|T_n|^s \le 2^{s-1} \left(E|\overline{T}_n|^s + \left| \sqrt{n} \int_a^{1-a} g_1(u) \left[E\widehat{F}_n(u) - F(u) \right] du \right| \right)^s \le$$

$$\le c_{23} \left(\int_a^{1-a} g(u)\, du \right)^{\frac{s}{2}} + \left| O\left(\sqrt{n}\, h^2\right) \int_a^{1-a} g(u)\, du \right|^s \le c_{24} \left(\int_a^{1-a} g(u)\, du \right)^{\frac{s}{2}}.$$

Lemma 2 is proved.

Let us introduce the following random processes:

$$\overline{T}_n(t) = \sqrt{n} \int_a^t \left(\widehat{F}_n(u) - E\widehat{F}_n(u) \right) \psi(F(u))\, du,$$

$$T_n(t) = \sqrt{n} \int_a^t \left(\widehat{F}_n(u) - F(u) \right) \psi(F(u))\, du.$$

Theorem 5. 1^0. *Let the conditions of the item* (a) *of Theorem* 4 *be fulfilled. Then for all continuous functionals* $f(\cdot)$ *on* $C[a, 1-a]$, *the distribution* $f(\overline{T}_n(t))$ *converges to the distribution* $f(W(t-a))$ *where* $W(t-a)$, $a \le t \le 1-a$, *is a Wiener process with a correlation function* $r(s,t) = \min(t-a, s-a)$, $W(t-a) = 0$, $t = a$.

2^0. *Let the conditions of the item* (b) *of Theorem* 4 *be fulfilled. Then for all continuous functionals* $f(\cdot)$ *on* $C[a, 1-a]$, *the distribution* $f(T_n(t))$ *converges to the distribution* $f(W(t-a))$.

Proof. First we will show that the finite-dimensional distributions of processes $\overline{T}_n(t)$ converge to the finite-dimensional distribution of a process $W(t-a)$, $t \ge a$. Let us consider one moment of time t_1. We have to show that

$$\overline{T}_n(t_1) \xrightarrow{d} W(t_1 - a). \tag{39}$$

To prove (39), it suffices to take $g(x) = I_{[a,t_1)}(x)$ in (25). Then, by virtue of Theorem 4,

$$\overline{T}_n(t_1) \xrightarrow{d} N\left(0, \int_a^{1-a} I_{[a,t_1)}(x)\, dx\right) = N(0, t_1 - a).$$

Let us now consider two moments of time t_1, t_2, $t_1 < t_2$. We have to show that

$$\left(\overline{T}_n(t_1), \overline{T}_n(t_2)\right) \xrightarrow{d} \left(W(t_1 - a), W(t_2 - a)\right). \tag{40}$$

To prove (40), it suffices to take in (25)

$$g(x) = (\lambda_1 + \lambda_2) I_{[a,t_1)}(x) + \lambda_2 I_{[t_1,t_2)}(x),$$

where λ_1 and λ_2 are arbitrary finite numbers. Then, by virtue of Theorem 4,

$$\lambda_1 \overline{T}_n(t_1) + \lambda_2 \overline{T}_n(t_2) \xrightarrow{d} N\left(0, (\lambda_1 + \lambda_2)^2 (t_1 - a) + \lambda_2^2 (t_2 - t_1)\right).$$

On the other hand,

$$\lambda_1 W(t_1 - a) + \lambda_2 W(t_2 - a) =$$

$$= (\lambda_1 + \lambda_2)\big[W(t_1 - a) - W(0)\big] + \lambda_2 \big[W(t_2 - a) - W(t_1 - a)\big]$$

is distributed as $N\left(0, (\lambda_1 + \lambda_2)^2 (t_1 - a) + \lambda_2^2 (t_2 - t_1)\right)$. Therefore (40) holds. The case of three and more number of moments is considered analogously. Therefore the finite-dimensional distributions of processes $\overline{T}_n(t)$ converge to the finite-dimensional distributions of a Wiener process $W(t-a)$, $a \le t \le 1-a$ with a correlation function

$$r(t_1, t_2) = \min(t_1 - a, t_2 - a), \quad W(t-a) = 0, \quad t = a.$$

Now we will show that the sequence $\{\overline{T}_n(t)\}$ is dense, i. e., the sequence of the corresponding distributions is dense. For this it suffices to show that for any $t_1, t_2 \in \in [a, 1-a]$ and all n

$$E\left|\overline{T}_n(t_1) - \overline{T}_n(t_2)\right|^s \le c_{25}|t_1 - t_2|^{\frac{s}{2}}, \quad s > 2.$$

Indeed, this inequality is obtained from (36) for $g(x) = I_{[t_1,t_2]}(x)$.

Further, taking (35), (37) and the statement b) of Theorem 4 into account, we easily ascertain that the finite-dimensional distributions of processes $T_n(t)$ converge to the finite-dimensional distributions of a Wiener process $W(t-a)$, and also that

$$E\left|T_n(t_1) - T_n(t_2)\right|^s \le c_{26}|t_1 - t_2|^{\frac{s}{2}}, \quad s > 2.$$

Hence, from Theorem 2 of the monograph [3, p. 583] the proof of the theorem follows.

Application. By virtue of Theorem 5 and the Corollary of Theorem 1 from [3, p. 371] we can write that

$$P\left\{T_n^+ = \max_{a \le t \le 1-a} T_n(t) > \lambda\right\} \longrightarrow$$

$$\longrightarrow G(\lambda) = \frac{2}{\sqrt{2\pi(1-2a)}} \int\limits_{\lambda}^{\infty} \exp\left\{-\frac{x^2}{2(1-2a)}\right\} dx$$

$\left(a \text{ is a prescribed number, } 0 < a < \dfrac{1}{2}\right)$ as $n \to \infty$.

This result makes it possible to construct tests of a level α, $0 < \alpha < 1$, for testing the hypothesis H_0 by which

$$H_0\colon \lim_{n\to\infty} E\widehat{F}_n(x) = F_0(x), \quad a \le x \le 1-a,$$

in the alternative hypothesis

$$H_1\colon \int\limits_a^{1-a} \psi(F_0(x))\left(\lim_{n\to\infty} E\widehat{F}_n(x) - F_0(x)\right) dx > 0.$$

Let λ_α be the critical value, $G(\lambda_\alpha) = \alpha$. If as a result of the experiment it turns out that $T_n^+ \ge \lambda_\alpha$, then the hypothesis H_0 must be rejected.

***Remark* 3.** Let t_i be the partitioning points of an interval $[0, c_F]$, $c_F = \inf\{x \ge \ge 0\colon F(x) = 1\} < \infty$, chosen from the relation $H(t_j) = \dfrac{2j-1}{2n}$, $j = \overline{1, n}$, where

$$H(x) = \int\limits_0^x h(u)\, du,$$

$h(u)$ is some known density of a distribution on $[0, c_F]$ and $h(x) \ge \mu > 0$ for all $x \in [0, c_F]$. In that case, by a reasoning analogous to that used above we can obtain a generalization of the results of the present study.

***Remark* 4.** Some ideas of the proof of Theorem 4 are borrowed from the interesting paper by A. V. Ivanov [5].

1. *Mandzhgaladze K. V.* On an estimator of a distribution function and its moments // Soobshch. Akad. Nauk Gruz. SSR. – 1986. – **124**, № 2. – S. 261 – 263.
2. *Parzen E.* On estimation of a probability density function and mode // Ann. Math. Statist.– 1962. – **33**. – P. 1065 – 1076.
3. *Whittle P.* Bounds for the moments of linear and quadratic forms in independent variables // Teor. Ver. i Prim. – 1960. – **5**. – S. 331 – 335.
4. *Gikhman I. I., Skorokhod A. V.* Introduction to the theory of random processes (in Russian). – Moskva: Gosudarstv. Izdat. Fiz.-Mat. Lit., 1965.
5. *Ivanov A. V.* Properties of a nonparametric estimate of the regression function (in Russian) // Dokl. Akad. Nauk Ukr. SSR. Ser. A. – 1979. – № 7. – S. 499 – 502, 589.

Received 23.07.10