A new Yang-type estimator of Distribution Function at Quantal Response over Indirect Data

MIKHAIL TIKHOV, MAXIM IVKIN
Department Applied Theory Probability
Nizhny Novgorod State University,
23 Prospekt Gagarina, 603950,
Nizhny Novgorod, RUSSIA
E-mail: tikhovm@mail.ru, ivkin_max@mail.ru

Abstract: – In the present work we consider the problem of the kernel estimation of the distribution function and the quantiles of this distribution in dose-effect relationship when the data are observed with error. We make both theoretical and computer research of estimations of the multivariate estimators of the distribution function in dose-effect relationship and in a case when measurements are made with an error having normal distribution. We offer also procedure of reduction of a measurement error.

Keywords: - Response model, multivariate data, nonparametric kernel estimates, measurement error.

1 Introduction

The problem of nonparametrically estimating of the distribution function and the quantiles of this distributions at quantal response, when the data are observed with an error, has attracted a great deal of interest [1]-[7]. One of many methods of nonparametric estimation is the kernel method [8]-[16]. Pros of the kernel method: easy to compute and always work. Basically estimations of one-dimensional distribution function a random variables were studied. It was supposed also that measurements are made without errors. In the present work we make both theoretical and computer research of estimations of multivariate estimators of the distribution function at quantal response and in case when measurements are made with an error having normal distribution. Also we offer the procedure of reduction of a measurement error.

2 Problem Formulation

Let $(\boldsymbol{U}_1,W_1),(\boldsymbol{U}_2,W_2),\dots,(\boldsymbol{U}_n,W_n)$ be independent, identically distributed (iid.) random $(d+1)$-vectors where $\{\boldsymbol{U}_i\}, 1\le i\le n$, is d-vectors with bounded continuous density $f(\boldsymbol{x})$, $W_i = I(\boldsymbol{X}_i < \boldsymbol{U}_i)$ is the indicator of an event $(\boldsymbol{X}_i < \boldsymbol{U}_i)$, d-vectors $\boldsymbol{X}_i$ has distribution function $Q(\boldsymbol{x}) = \mathbf{P}(\boldsymbol{X} < \boldsymbol{x})$ and continuous density $q(\boldsymbol{x}) > 0$. The problem is to estimate the distribution function $Q(\boldsymbol{x})$ from the sample $\mathsf{U}^{(n)} = \{(\boldsymbol{U}_i,W_i),\ 1\le i\le n\}$.

Usually as an estimate of $Q(\boldsymbol{x})$ nonparametric estimators are used.

In a case $d=1$ kernel estimators

$$\hat{Q}_n(x) = S_{2n}(x)/S_{1n}(x), \tag{1}$$

are applied, where $S_{jn}(x) = \frac{1}{nh}\sum_{i=1}^{n} W_i^{j-1} L\left(\frac{u_i - x}{h}\right)$, $j=1,2$ and the kernel $L(x)$ is nonnegative even function, and $\int L(x)\,dx = 1$. We have $\hat{F}_n(x)=0$ if $S_{1n}(x)=0$. As h we take $h=n^{-1/5}$.

S.S.Yang [13] proposed as regression function $Q(x) = \mathbf{E}(W \mid U = x)$ the statistic $Q_n^*(x)$ of the form

$$Q_n^*(x) = \frac{1}{nh}\sum_{i=1}^{n} L\left(\frac{i/n - F_n(x)}{h}\right) W_n^{[i]}, \tag{2}$$

where $F_n(x) = n^{-1}\sum_{i=1}^{n} I(U_i < x)$ is the empirical distribution function random variable (r.v.) U, identically distributed with rv's $U_1,U_2,\dots,U_n$. Let $U_n^{(1)} < \dots < U_n^{(i)} < \dots < U_n^{(n)}$ be order statistics, and $W_n^{[i]}$ pared with $U_n^{(i)}$ is called concomitant of the i-th order statistics in the sample $\mathsf{U}^{(n)}$.

Let $k = k(n)$ be a sequence of positive integers, and $\rho = \rho_n$ be the Euclidean distance between x and its k-th nearest neighbor. The nearest neighbor estimate is

$$\tilde{Q}_n(x) = T_{2n}(x)/T_{1n}(x), \tag{3}$$

where $T_{jn}(x) = \frac{1}{n\rho}\sum_{i=1}^{n} W_i^{j-1} L\left(\frac{u_i - x}{\rho}\right)$.

Let's notice that the estimate (2) is also a nearest neighbor estimate, but now neighbors are defined in terms of distance based on the empirical distribution function.

The estimate $\hat{Q}_n(x)$ has variance $\sigma_1^2 = (Q(x)(1-Q(x))\|L\|^2 / f(x))(1+o(1/(nh))$ (see, [16]), where $\| L \|^2 = \int L^2(x)\,dx$, therefore, if the density $f(x)=0$, then this case is better to use the estimate (2).

Consider an estimate of $Q(\boldsymbol{x})$ given by

$$\hat{Q}_n(\boldsymbol{x}) = \frac{T_{2n}(\boldsymbol{x})}{T_{1n}(\boldsymbol{x})}, \tag{4}$$

where

$$T_{2n}(\boldsymbol{x}) = \frac{1}{n\rho^d}\sum_{j=1}^{n} W_j L\left(\frac{\boldsymbol{U}_j - \boldsymbol{x}}{\rho}\right) = \frac{1}{n}\sum_{j=1}^{n} W_j L_\rho(\boldsymbol{U}_j - \boldsymbol{x}),$$

$$T_{1n}(\boldsymbol{x}) = \frac{1}{n\rho^d}\sum_{j=1}^{n} L\left(\frac{\boldsymbol{U}_j - \boldsymbol{x}}{\rho}\right) = \frac{1}{n}\sum_{j=1}^{n} L_\rho(\boldsymbol{U}_j - \boldsymbol{x}),$$

$\rho = \rho_n(\boldsymbol{x})$ is the Euclidean distance between $\boldsymbol{x}$ and kth nearest neighbor of $\boldsymbol{x}$ among the $\boldsymbol{U}_j$'s, $L(\boldsymbol{x})$ is a bounded integrable weight function with $\int L(\boldsymbol{u})\,d\boldsymbol{u} = 1$, $k = k(n)$ is a sequence of positive integer such that $k\to\infty,\ k/n\to 0$ as $n\to\infty.$

However the real data which we observe have measurement errors [3], [17]. Data measured with errors occur frequently in many scientific fields. Ignoring measurement error can bring forth biased estimates and lead to erroneous conclusions to various degrees in a data analysis. We will assume that the data measurements are made with an error, which is a random variable with a known or an unknown continuous distribution function $G(x)$ and density $g(x)$. In other words, instead of the data $(\boldsymbol{U}_i, W_i), 1\le i\le n,$ (*direct data*) we observe $(\boldsymbol{Y}_i, W_i), 1\le i\le n,$ where $\boldsymbol{Y}_i = \boldsymbol{U}_i + \boldsymbol{\varepsilon}_i,\quad W_i = I(\boldsymbol{X}_i < \boldsymbol{U}_i)$ (*indirect data*). How can limit distributions of estimator $\hat{Q}_n(\boldsymbol{x})$ change? In this paper we consider also the limit distributions of the estimators

$$\hat{Q}_n(\boldsymbol{x}) = \frac{T_{2n}(\boldsymbol{x})}{T_{1n}(\boldsymbol{x})}, \tag{5}$$

where

$$T_{2n}(\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^{n} W_j L_\rho(\boldsymbol{Y}_j - \boldsymbol{x}),\ T_{1n}(\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^{n} L_\rho(\boldsymbol{Y}_j - \boldsymbol{x}).$$

3 Preliminary Results

3.1. Asymptotic Normality of Linear Functions of Concomitants of Order Statistics

At the beginning we will quote results of work [8]. Let $(\boldsymbol{X}_i^\top, Y_i)^\top$ be a $\boldsymbol{R}^{N+1}$-valued time-series process on the probability space $(\Omega, \mathsf{A}, \mathbf{P})$. Let $Y_n^{(1)} < \ldots < Y_n^{(i)} < \ldots < Y_n^{(n)}$ be the order statistics; and $\boldsymbol{X}_n^{[n]}$ paired with $Y_n^{(i)}$ is called the concomitant of the i-th order statistics in the sample $\{\boldsymbol{X}_n^{[i]}, Y_n^{(i)}\}_{i=1}^n$.

In [8] it is prove the $\sqrt{n}$-asymptotic normality, under fairly mild regularity conditionals sush as

$$A(F_n) = \frac{1}{n}\sum_{i=1}^{n} J(i/n)\, b(\boldsymbol{X}_n^{[i]}, Y_n^{(i)}) =$$

$$= \int\limits_{\boldsymbol{R}^{N+1}} J(F_n(y))\, b(\boldsymbol{x}, y)\, dF_n(\boldsymbol{x}, y) \tag{6}$$

and

$$\hat{g}(Y_n(\alpha);\beta) = \frac{1}{(n-1)h_n}\sum_{s=1}^{n-1} [Y_n^{*[s]} - W_n^{[s]}]\times$$

$$\times K\left(\frac{F_n(Y_n(\alpha) - s/(n-1))}{h_n}\right), \tag{7}$$

where $J(\cdot)$ is a bounded smooth score function and $b(\boldsymbol{x}, y)$ is some $\boldsymbol{R}$-valued known function of $(\boldsymbol{x}^\top, y)^\top \in \boldsymbol{R}^{N+1}$, $F_T(\cdot)$ is the empirical distribution function; $K(\cdot)$ is a kernel (weight) function; $Y_i^* = g(\boldsymbol{Z}_i^{*\top}\boldsymbol{\alpha}) + \boldsymbol{X}_i^{*\top}\boldsymbol{\beta} + \varepsilon_i$, $\boldsymbol{Z}_i^{*\top}$ and $\boldsymbol{X}_i^{*\top}$ are random covariate vectors; $g(\cdot)$ represents an unknown, possibly non-differentiable, function; ε denote i.i.d. mean-zero random errors; $Y_t(\alpha) = \boldsymbol{Z}_t^{*\top}\boldsymbol{\alpha}$, $W_t(\beta) = \boldsymbol{X}_t^{*\top}\boldsymbol{\beta}$; and $\boldsymbol{X}_n^{[s]}(\beta) = (Y_n^{[s]}, W_n^{[s]}(\beta))$ denotes a vector of the concomitant of the order statistics $Y_n^{(s)}(\alpha)$ in the sample $\{(Y_1^*, W_1(\beta), Y_1(\alpha)), \ldots,$ $(Y_{t-1}^*, W_{t-1}(\beta), Y_{t-1}(\alpha)),\ (Y_{t+1}^*, W_{t+1}(\beta), Y_{t+1}(\alpha)), \ldots,$ $(Y_n^*, W_n(\beta), Y_n(\alpha))\}$.

Stute [22] shows that in the i.i.d. case, the asymptotic behavior of $\hat{g}(y;\beta)$ is the same as that of

$$\hat{g}^*(y;\beta) = \frac{1}{n-1}\sum_{s=1}^{n-1} J(s/(n-1))\{Y_n^{*[s]} - W_n^{[s]}\}, \tag{8}$$

where

$J_n(s/(n-1)) = h_n^{-1} K((F(y) - s/(n-1))/h_n)$.

Let T be a measure-preserving, I denote the Borel algebra of invariant sets $A\in\mathsf{A}$ such that $\mathsf{T}^{-1}A = A$. Let the quantity $\|A\|_p$ is the L_p-norm of A, i.e. $\{\mathbf{E}[|A|^p]\}^{1/p}$; $\|A\|_{p,\mathsf{I}}$ is the L_p-norm of A condotional on I, i.e. $\{\mathbf{E}[|A|^p \,|\, \mathsf{I}\,]\}^{1/p}$.

Let

$$m_b(y;\mathrm{I}) = \mathbf{E}(b(\boldsymbol{X},Y) \mid Y = y, \mathrm{I}),$$
$$m_b(y) = \mathbf{E}(b(\boldsymbol{X},Y) \mid Y = y).$$

The following regularity conditions are introduced in [8].

C1 Moment Bounds:

For a given integer, $p>1$,

$$\max\{\| b(\boldsymbol{X}_1,Y_1)\|_{p,\mathrm{I}}, \| b(\boldsymbol{X}_1,Y_1)\|_{2p/(p-1),\mathrm{I}}\} < \infty.$$

C2 Conditional Moments:

(a) $\| m_b'(Y_1;\mathrm{I})\|_{p^*,\mathrm{I}} < \infty$, where $m_b'(\cdot;\mathrm{I})$ is the first derivative of $m_b(\cdot;\mathrm{I})$.

(b)

$$\lim_{T\to\infty} \left\| \sup_y |\mathbf{P}(Y_T \le y \mid \mathcal{F}_{Y_0}) - F(y \mid \mathrm{I}) \right\|_{q^*,\mathrm{I}} = 0,$$

where p^* and q^* are such integers that

$$1/p^* + 1/q^* = 1 + (p-1)/p.$$

C3 Conditional Joint Moments:

(a)

$$\sum_{t=1}^{\infty} \left\| \| m_b(Y_t;\mathsf{F}_{Y_0}) - m_b(Y_t;\mathrm{I}) \|_{2,\mathsf{F}_{Y_0}} \right\|_{p/(p-1),\mathrm{I}} < \infty$$

(b)

$$\sum_{t=1}^{\infty} \Big\| \mathbf{E}(b(\boldsymbol{X}_t,Y_t) b(\boldsymbol{X}_0,Y_0) \mid \mathsf{F}_{Y_t}, \mathrm{I}) - m_b(Y_t;\mathrm{I}) m_b(Y_0;\mathrm{I}) \Big\|_{p/(p-1),\mathrm{I}} < \infty.$$

Theorem 1 [8]**.** *Suppose that Assumptions* **C1**,**C2**, *and* **C3** *hold. Then*

$$\sqrt{n}(A(F_n) - A(F)) \xrightarrow[n\to\infty]{d} N(0, \sigma_W^2(\mathrm{I})),$$

where $\sigma_W^2(\mathrm{I}) = \mathbf{E}(W_1^{*2} \mid \mathrm{I}) = \mathbf{E}(W_0^2 \mid \mathrm{I})$.

3.2. Distribution of *k*-Nearest Neighbour Distances

Let us first consider the probability density $p(\boldsymbol{x})$ of the distance ρ between $\boldsymbol{x}$ and the *k*th nearest neighbor $\boldsymbol{x}$. Let $S_r = \{\boldsymbol{z} : \|\boldsymbol{z} - \boldsymbol{x}\| < r\}$, $G(r) = \mathbf{P}(S_r)$, and

$$G'(r) = \lim_{\delta\to 0} \frac{1}{\delta}\left[\int_{S_{r+\delta}} f(\boldsymbol{t})\,d\boldsymbol{t} - \int_{S_r} f(\boldsymbol{t})\,d\boldsymbol{t}\right] = \int_{\|\boldsymbol{x}-\boldsymbol{t}\| = r} f(\boldsymbol{t})\,d\sigma(\boldsymbol{t}), \qquad (9)$$

where $\mathbf{P}$ is the probability measure with density f, and σ is the surface area of the sphere $\|\boldsymbol{x} - \boldsymbol{t}\| = r$,

$$c_d = \frac{\pi^{d/2} r^d}{\Gamma((d+2)/2)}, \ \beta_d = d \cdot c_d.$$

Thus the density of ρ is

$$p_n(r) = \frac{n!}{(k-1)!(n-k)!} G^{k-1}(r)(1-G(r))^{n-k} G'(r). \quad (10)$$

The joint density of the *k*th nearest neighbor $\boldsymbol{h}$, the $k-1$ observations $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_{k-1}$ falling within the sphere about $\boldsymbol{x}$ whose radius is determined by $\boldsymbol{x}$ and $\boldsymbol{h}$, and the remaining $(n-k)$ observations $\boldsymbol{V}_1, \boldsymbol{V}_2, ..., \boldsymbol{V}_{n-k}$ falling outside this sphere, is given as following. Consider the joint density of $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$. There are n choices possible for $\boldsymbol{h}$. Given that $\boldsymbol{h}$ is chosen, there are $\binom{n-1}{k-1}$ possible choices for the $k-1$ observations falling within the sphere (and this determines the $n-k$ falling outside the sphere). The joint density of $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_{k-1}$, $\boldsymbol{V}_1, \boldsymbol{V}_2, ..., \boldsymbol{V}_{n-k}$ and $\boldsymbol{h}$ is then (see, [15],[19])

$$p(\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_{k-1}; \boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_{n-k}; \boldsymbol{h} | r) = n\binom{n-1}{k-1}\left\{\prod_{j=1}^{k-1} f(\boldsymbol{y}_j)\delta(\boldsymbol{y}_j, S_r)\right\} \times \left\{\prod_{l=1}^{n-k} f(\boldsymbol{v}_l)\delta(\boldsymbol{v}_l, (\overline{S}_r)^c)\right\} f(\boldsymbol{h}), \qquad (11)$$

where

$$\delta(\boldsymbol{z}, A) = \begin{cases} 1 \text{ if } \boldsymbol{z} \in A, \\ 0 \text{ otherwise,} \end{cases}$$

$(\overline{S}_r)^c$ is the complement of $\overline{S}_r$. This implies that the conditional distribution of the the $\boldsymbol{Y}_j$'s, $\boldsymbol{V}_l$'s and h given $\rho = r$ is

$$p(\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_{k-1}; \boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_{n-k}; \boldsymbol{h} | r) = \prod_{j=1}^{k-1}\left(\frac{f(\boldsymbol{y}_j)}{G(r)}\right)\prod_{l=1}^{n-k}\left(\frac{f(\boldsymbol{v}_l)}{1-G(r)}\right)\frac{f(\boldsymbol{h})}{G'(r)}, \quad (12)$$

so that the $\boldsymbol{Y}_j$'s, $\boldsymbol{V}_l$'s and $\boldsymbol{h}$ are conditionally independent given $\rho = r$ with respective marginal densities

$$\frac{f(\boldsymbol{y}_j)}{G(r)}, \ \frac{f(\boldsymbol{v}_l)}{1-G(r)}, \text{ and } \frac{f(\boldsymbol{h})}{G'(r)},$$

$$\{\boldsymbol{y} : \|\boldsymbol{x} - \boldsymbol{y}\| < r\}, \ \{\boldsymbol{v} : \|\boldsymbol{x} - \boldsymbol{v}\| > r\}, \ \{\boldsymbol{h} : \|\boldsymbol{x} - \boldsymbol{h}\| = r\}$$

where the conditional density of $\boldsymbol{h}$ given ρ is to be integrated with respect to the surface measure on the sphere of radius r about $\boldsymbol{x}$.

We are interested in computing moments of various functions of ρ. It is clear from what has been stated above that ρ has the same distribution as $G^{-1}(\xi)$, where ξ is the *k*th order statistic from an i.i.d. uniform $(0,1)$ sample of size n. If we just assume f is bounded and continuous we have

$$G(r) = \int_{B_r} f(\boldsymbol{u})\,d\boldsymbol{u} = c_d f(\boldsymbol{x}) r^d +$$

$$+\int_{B_r} (f(\boldsymbol{u}) - f(\boldsymbol{x}))\,d\boldsymbol{u} = c_d f(\boldsymbol{x}) r^d + o(r^d)$$

as $r \downarrow 0$.

Then $t = G(r)$ it follows that when $f(\boldsymbol{x}) > 0$

$$(G^{-1}(t))^{\lambda} = r^{\lambda} = \left(\frac{t}{c_d f(\boldsymbol{x})}\right)^{\lambda/d} + o(t^{\lambda/d}) . \quad (13)$$

In addition, from Theorem 1 [17] we have

$$G^{-1}(z) = \left[\frac{z}{c_0 f(\boldsymbol{x})}\right]^{1/d} -$$

$$-\left\{\frac{c_2 \nabla^2 f(\boldsymbol{x})}{2 p c_0 f(\boldsymbol{x})}\right\} \cdot \left\{\frac{z}{c_0 f(\boldsymbol{x})}\right\}^{3/d} + o(z^{3/d}) \quad (14)$$

under conditions

$$\int \|\boldsymbol{y}\|^2 L(\boldsymbol{y})\,d\boldsymbol{y} < \infty, \int L(\boldsymbol{y})\,d\boldsymbol{y} = 1,$$

$$\int y_i L(\boldsymbol{y})\,d\boldsymbol{y} = \int y_i y_j L(\boldsymbol{y})\,d\boldsymbol{y} = 0,\ i \neq j,$$

$$\int y_i^2 L(\boldsymbol{y})\,d\boldsymbol{y} > 0 \text{ for any } i .$$

4 Main Results

4.1. Direct Data

Let's consider a difference $\hat{Q}_n(\boldsymbol{x}) - Q(\boldsymbol{x})$.

We have (see [16])

$$\tau_n(\boldsymbol{x}) =$$

$$= \frac{T_{2n}(\boldsymbol{x})(f(\boldsymbol{x}) - T_{1n}(\boldsymbol{x}))}{T_{1n}(\boldsymbol{x}) f(\boldsymbol{x})} + \frac{(T_{2n}(\boldsymbol{x}) - Q(\boldsymbol{x}) f(\boldsymbol{x}))}{f(\boldsymbol{x})} . \quad (15)$$

If we show that $T_{1n}(\boldsymbol{x}) - f(\boldsymbol{x}) \xrightarrow[n\to\infty]{p} 0$ and $T_{2n}(\boldsymbol{x}) - Q(\boldsymbol{x}) f(\boldsymbol{x}) \xrightarrow[n\to\infty]{p} 0$, then from Slutsky's theorem (see [18], p.388, Theorem A.102), owing to boundedness of $Q(\boldsymbol{x}) f(\boldsymbol{x})$ and using that $f(\boldsymbol{x}) \geq c_0 > 0$ we obtain the convergence $\tau_n(\boldsymbol{x})$ in probability to zero as $n \to \infty$: $\tau_n(\boldsymbol{x}) \xrightarrow[n\to\infty]{p} 0$ for every fixed $\boldsymbol{x}$. Besides, from this relation we receive the limiting distribution of $\tau_n(\boldsymbol{x})$.

Let's consider the characteristic function $\varphi_{1n}(\theta)$ of the statistic $T_{1n}(\boldsymbol{x})$. Let

$$\varphi_{1n}(\theta) = \mathbf{E}\left(\exp\left(i\theta T_{1n}(\boldsymbol{x})\right)\right).$$

From (3) we derive

$$\varphi_{1n}(\theta) = \int \left(\psi_{1n}(\theta, r)\right)^{k-1} \psi_{2n}(\theta, r) \left(\psi_{3n}(\theta, r)\right)^{n-k} p_n(r)\,dr, \quad (16)$$

where

$$\psi_{1n}(\theta, r) = \int_{\|\boldsymbol{y}-\boldsymbol{x}\|<r} \exp\left(\frac{i\theta}{n} Q(\boldsymbol{y}) L_r(\boldsymbol{y}-\boldsymbol{x})\right) \frac{f(\boldsymbol{y})}{G(r)}\,d\boldsymbol{y}, \quad (17)$$

$$\psi_{2n}(\theta, r) = \int_{\|\boldsymbol{t}-\boldsymbol{x}\|=r} \exp\left(\frac{i\theta}{n} Q(\boldsymbol{t}) L_r(\boldsymbol{t}-\boldsymbol{x})\right) \frac{f(\boldsymbol{t})}{1-G(r)}\,d\boldsymbol{t} \quad (18)$$

$$\psi_{3n}(\theta, r) = \int_{\|\boldsymbol{v}-\boldsymbol{x}\|>r} \exp\left(\frac{i\theta}{n} Q(\boldsymbol{v}) L_r(\boldsymbol{v}-\boldsymbol{x})\right) \frac{f(\boldsymbol{v})}{1-G(r)}\,d\boldsymbol{v}. \quad (19)$$

By (5), the first term is equals

$$\psi_{1n}(\theta, r) = \frac{1}{G(r)} \int_{\|\boldsymbol{u}\|<r} \exp\left(\frac{i\theta}{n} L_r(\boldsymbol{u})\right) f(\boldsymbol{x} - \boldsymbol{u} r)\,d\boldsymbol{u}. \quad (20)$$

Lemma 1 [16]. *For every $\alpha \in \mathbf{R}^1$ and $n \geq 0$,*

$$\left| e^{i\alpha} - \sum_{k=0}^{n} \frac{(i\alpha)^k}{k!} \right| \leq \min\left\{ \frac{|\alpha|^{n+1}}{(n+1)!}, \frac{2|\alpha|^n}{n!} \right\}. \quad (21)$$

Now, in virtue of the condition (4) on the kernel $L(\boldsymbol{x})$, using the fact that $L(\boldsymbol{x}) \leq M_1$, $f(\boldsymbol{x}) \leq M_2$, applying the results of the Lemma 1, we conclude that

$$\left| \int_{\|\boldsymbol{u}\|<r} \left(\exp\left(\frac{i\theta}{n} L_r(\boldsymbol{x}-\boldsymbol{u})\right) - 1 - \frac{1}{n}\left(i\theta L_r(\boldsymbol{x}-\boldsymbol{u})\right) - \right.\right.$$

$$\left.\left. - \frac{1}{2n^2}\left(i\theta L_r(\boldsymbol{x}-\boldsymbol{u})\right)^2 \right) f(\boldsymbol{u})\,d\boldsymbol{u} \right| \leq$$

$$\leq \frac{1}{6n^3} \int_{\|\boldsymbol{u}\|<r} |\theta|^3 \left(L_r(\boldsymbol{x}-\boldsymbol{u})\right)^3 f(\boldsymbol{u})\,d\boldsymbol{u} =$$

$$= \frac{|\theta|^3}{6n^3 r^{2p}} \int_{\|\boldsymbol{u}\|<1} L^3(\boldsymbol{u}) f(\boldsymbol{x} - \boldsymbol{u} r)\,d\boldsymbol{u} \leq$$

$$\frac{|\theta|^3 M_1^3 M_2}{6n^3 r^{2d}} = \frac{|\theta|^3 M_1^3 M_2 c_d f^2(\boldsymbol{x})}{6nk^2} . \quad (22)$$

Hence,

$$\int_{\|\boldsymbol{u}\|<r} \left(\exp\left(\frac{i\theta}{n} L_r(\boldsymbol{x}-\boldsymbol{u})\right) - 1 - \frac{1}{n}\left(i\theta L_r(\boldsymbol{x}-\boldsymbol{u})\right) - \right.$$

$$\left. - \frac{1}{2n^2}\left(i\theta L_r(\boldsymbol{x}-\boldsymbol{u})\right)^2 \right) f(\boldsymbol{u})\,d\boldsymbol{u} = O(k^{-3}). \quad (23)$$

Further,

$$\int_{\|\boldsymbol{t}\|=r} \left| \exp\left(\frac{i\theta}{n} L_r(\boldsymbol{t})\right) - 1 \right| \frac{f(\boldsymbol{x}+\boldsymbol{t})}{G'(r)}\,d\boldsymbol{t} \leq$$

$$\leq \frac{|\theta|}{n} \int_{\|\boldsymbol{u}\|=1} L(\boldsymbol{u}) \frac{f(\boldsymbol{x}+\boldsymbol{u} r)}{G'(r)}\,d\boldsymbol{u} = O\left(\frac{1}{n}\right), \quad (24)$$

thus $\psi_{2n}(\theta, r) \to 1$ as $n \to \infty$.

Similarly,

$$\left| \int_{\|\boldsymbol{u}\|>r} \left(\exp\left(\frac{i\theta}{n} L_r(\boldsymbol{x}-\boldsymbol{u})\right) - 1 - \right.\right.$$

$$\left.\left. - \frac{1}{n}\left(i\theta L_r(\boldsymbol{x}-\boldsymbol{u})\right) \right) f(\boldsymbol{u})\,d\boldsymbol{u} \right| \leq$$

$$\leq \frac{\theta^2}{2n^2} \int_{\|\boldsymbol{u}\|>r} \left(L_r(\boldsymbol{u}-\boldsymbol{x})\right)^2 f(\boldsymbol{u})\,d\boldsymbol{u} =$$

$$=\frac{\theta^2}{2n^2r^d}\int_{\|\boldsymbol{u}\|>1}\left(L(\boldsymbol{u})\right)^2 f(\boldsymbol{x}-\boldsymbol{u}r)\,d\boldsymbol{u}\le\frac{\theta^2M_1^2M_2}{2n^2r^d}=$$
$$=\frac{\theta^2M_1^2M_2c_d f(\boldsymbol{x})}{2nk}=O\left(\frac{1}{nk}\right)=o\left(\frac{1}{n}\right). \quad (25)$$

For further we use the results of works [15] under the conditions of Theorem 1

$$\mathbf{E}(T_{1n}(\boldsymbol{x}))=f(\boldsymbol{x})+\frac{a_d}{(f(\boldsymbol{x}))^{2/d}}P(f)(\boldsymbol{x})\left(\frac{k}{n}\right)^{2/d}+$$
$$+\frac{c_d f(\boldsymbol{x})}{k}\int_{\|\boldsymbol{u}\|=1}L(\boldsymbol{u})\,d\Sigma_0+o\left(\left(\frac{k}{n}\right)^{2/d}+\frac{1}{k}\right), \quad (26)$$

where $P(f)(\boldsymbol{x})=\sum_{i,j}\int u_iu_jL(\boldsymbol{u})\,d\boldsymbol{u}\,D_iD_jf(\boldsymbol{x}),$

$H_f(\boldsymbol{x})=(D_iD_jf(\boldsymbol{x}))-$ Hessian matrix,

$$\mathbf{D}(T_{1n}(\boldsymbol{x}))=\frac{c_d f^2(\boldsymbol{x})}{k}\int L^2(\boldsymbol{u})\,d\boldsymbol{u}+o\left(\frac{1}{k}\right), \quad (27)$$

$a_d=(\Gamma((d+2)/2))^{2/d}/(2\pi)$,

and, accordingly,

$$\frac{1}{n}\int_{\|\boldsymbol{v}\|>r}L_r^2(\boldsymbol{v}-\boldsymbol{x})\frac{f(\boldsymbol{v})}{1-G(r)}d\boldsymbol{v}=$$
$$=\frac{c_d f^2(\boldsymbol{x})}{k}\int L^2(\boldsymbol{u})\,d\boldsymbol{u}+o\left(\frac{1}{k}\right). \quad (28)$$

Therefore uniformly in $\theta\in[-T,T]$, where T is a real number

$$\int\exp\left(i\theta n^{-1}L_r(\boldsymbol{u}-\boldsymbol{x})\right)f(\boldsymbol{u})\,d\boldsymbol{u}=$$
$$=1+i\theta f(\boldsymbol{x})k^{-1}-(1/2)\theta^2c_d f^2(\boldsymbol{x})\|L\|^2k^{-1}+$$
$$+o\left(k^{-1}\right),\left(n\to\infty\right);\ \|L\|^2=\int L^2(\boldsymbol{u})\,d\boldsymbol{u}. \quad (29)$$

Decompose $\ln\varphi_{1n}(t)$ in a series on exponents $i\alpha-\beta=i\theta\frac{f(\boldsymbol{x})}{k}-\frac{\theta^2}{2}\frac{c_d f^2(\boldsymbol{x})\|L\|^2}{k}$ to the second term and use the fact that for any real α,

$$\left|\ln(1+i\alpha)-i\alpha\right|\le\frac{\alpha^2}{2}\ \text{(see [15])}. \quad (30)$$

Let $\alpha=\theta\frac{f(\boldsymbol{x})}{k}$ and $\beta=\frac{\theta^2}{2}\frac{c_d f^2(\boldsymbol{x})\|L\|^2}{k}$.

Then

$$\ln\left(1+i\alpha-\beta\right)=\ln\left(1-\beta\right)+\ln\left(1+\frac{i\alpha}{1-\beta}\right)=$$
$$=-\beta+\frac{i\alpha}{1-\beta}+O\left(k^{-2}\right)=-\beta+i\alpha+O\left(k^{-2}\right), \quad (31)$$

since

$$\beta^2=O\left(k^{-2}\right),\ \frac{\alpha^2}{\left(1-\beta\right)^2}\le\alpha^2=O\left(k^{-2}\right),\ \text{as } |\beta|<\frac{1}{2}.$$

Hence

$$\left|\ln\left(1+i\alpha-\beta\right)-i\alpha+\beta\right|=O\left(k^{-2}\right). \quad (32)$$

Therefore, as $n\to\infty$,

$$\left|\ln\left(1+i\theta f(\boldsymbol{x})k^{-1}-(1/2)\theta^2c_d f^2(\boldsymbol{x})\|L\|^2k^{-1}\right)-\right.$$
$$\left.-\left(i\theta f(\boldsymbol{x})k^{-1}-(1/2)\theta^2c_d f^2(\boldsymbol{x})\|L\|^2k^{-1}\right)\right|=$$
$$=O\left(k^{-2}\right)=o\left(k^{-1}\right), \quad (33)$$

since a function $f(\boldsymbol{x})$ for a fixed $\boldsymbol{x}$ is bounded, and $\|L\|^2<\infty$, $\theta\in[-T,T]$, then $o(1)$ converges uniformly to zero.

Since $b_{1n}\,p_n(b_{1n}r+b_{0n})$, where $b_{1n},b_{0n}-$ appropriate normalizing multipliers, converges uniformly on any bounded interval $\left(-C,C\right)$, $C>0$ to the density of the limit distribution (see [20]), and, given that the probability of hitting $b_{1n}^{-1}(\rho-b_{0n})$ into intervals $\left(-\infty,-C\right],\left[C,\infty\right)$ tend to zero (see [18]), and the function $\exp\left(i\theta f(\boldsymbol{x})k^{-1}-(1/2)\theta^2c_d f^2(\boldsymbol{x})\|L\|^2k^{-1}\right)$ is bounded, we have that

$$\mathbf{E}(\exp(it\sqrt{k}(T_{1n}(\boldsymbol{x})-f(\boldsymbol{x}))))\to$$
$$\to\exp\left(-(1/2)\theta^2c_d f^2(\boldsymbol{x})\|L\|^2\right) \quad (34)$$

as $n\to\infty$. Now from convergence of characteristic functions follows, that

$$\sqrt{k}(T_{1n}(\boldsymbol{x})-f(\boldsymbol{x}))\xrightarrow[n\to\infty]{d}N(0,c_d f^2(\boldsymbol{x})\|L\|^2),$$

Whence also follows that

$$T_{1n}(\boldsymbol{x})-f(\boldsymbol{x})\xrightarrow[n\to\infty]{p}0. \quad (35)$$

We now show that

$$\sqrt{k}(T_{2n}(\boldsymbol{x})-Q(\boldsymbol{x})f(\boldsymbol{x}))\xrightarrow[n\to\infty]{d}$$
$$\to N(0,c_dQ^2(\boldsymbol{x})f^2(\boldsymbol{x})\|L\|^2). \quad (36)$$

Let

$$\varphi_{2n}(\theta)=\mathbf{E}(\exp(i\theta T_{2n}(\boldsymbol{x})))=$$
$$=\mathbf{E}\left(\exp\left(i\theta n^{-1}\sum_{j=1}^{n}I(\boldsymbol{X}_j<\boldsymbol{U}_j)L_\rho(\boldsymbol{U}_j-\boldsymbol{x})\right)\right).$$

Passing on first to the conditional expectation provided $\boldsymbol{U}_j$, and then arguing as above with respect to the characteristic function of the statistics $T_{1n}(\boldsymbol{x})$, we obtain the following representation

$$\varphi_{2n}(\theta)=\int\left(\lambda_{1n}(\theta,r)\right)^{k-1}\lambda_{2n}(\theta,r)\left(\lambda_{3n}(\theta,r)\right)^{n-k}p_n(r)\,dr, \quad (37)$$

where

$$\lambda_{1n}(\theta,r)=\int_{\|\boldsymbol{y}-\boldsymbol{x}\|<r}\exp\left(\frac{i\theta}{n}Q(\boldsymbol{y})K_r(\boldsymbol{y}-\boldsymbol{x})\right)\frac{f(\boldsymbol{y})}{G(r)}d\boldsymbol{y},$$

(38)

$$\lambda_{2n}(\theta,r)=\int_{\|t-x\|=r}\exp\left(\frac{i\theta}{n}Q(t)K_r(t-x)\right)\frac{f(t)}{G'(r)}dt, \quad (39)$$

$$\lambda_{3n}(\theta,r)=\int_{\|v-x\|>r}\exp\left(\frac{i\theta}{n}Q(v)K_r(v-x)\right)\frac{f(v)}{1-G(r)}dv. \quad (40)$$

Repeating the above arguments, but for functions $\lambda_{1n}(\theta,r)$, $\lambda_{2n}(\theta,r)$, $\lambda_{3n}(\theta,r)$ and $\varphi_{2n}(\theta)$, we find that

$$\sqrt{k}(T_{2n}(x)-Q(x)f(x))\underset{n\to\infty}{\overset{d}{\to}}N(0,c_dQ^2(x)f^2(x)\|K\|^2). \quad (41)$$

Thus we have the following result.

Theorem 2. *Let the density be bounded and there is a third continuous bounded partial derivatives $f(x)$ and $Q(x)$, $\int\|u\|^2L(u)\,du<\infty$, $\int uL(u)\,du=\mathbf{0}$. Then*

(i) $\sqrt{k}\left(T_{1n}(x)-f(x)\right)\underset{n\to\infty}{\overset{d}{\to}}N(0,c_d f^2(x)\|L\|^2)$, (42)

(ii) $\sqrt{k}\left(T_{2n}(x)-Q(x)f(x)\right)\underset{n\to\infty}{\overset{d}{\to}}$

$$\to N(0,c_p f^2(x)Q^2(x)\|L\|^2). \quad (43)$$

The following theorem establishes the asymptotic normality of the estimator $\tilde{F}_n(x)$ of the distribution function $Q(x)$.

Theorem 3. *Let the conditions of Theorem 1 hold. Then*

$$\sqrt{k}\left(\hat{Q}_n(x)-Q(x)\right)\underset{n\to\infty}{\overset{d}{\to}}N(0,Q(x)(1-Q(x))\|L\|^2). \quad (44)$$

Proof. Let $T_1=T_{1n}(x)$, $T_2=T_{2n}(x)$, $Qf=Qf(x)=Q(x)f(x)$, $f=f(x)$.

We have [9]:

$$\frac{T_2}{T_1}-\frac{Qf}{f}=\frac{T_2-Qf}{f}-\frac{Qf}{f^2}(T_1-f)+ +O_p\left(\frac{(T_2-Qf)(T_1-f)}{f^2}\right)+O_p\left(\frac{Qf(T_1-f)^2}{f^3}\right).$$

Arguing as in [20] with respect to the statistics $T_{1n}(x)$ and $T_{2n}(x)$, it can be shown that with probability 1

$$\overline{\lim_{n\to\infty}}\sup_{\frac{c\ln n}{n}\le h\le 1}\frac{\sqrt{nh}\left\|T_1-\mathbf{E}(f)\right\|_\infty}{\sqrt{\max\left(\ln(1/h),\ln\ln n\right)}}=k_1(c)<\infty, \quad (45)$$

$$\overline{\lim_{n\to\infty}}\sup_{\frac{c\ln n}{n}\le h\le 1}\frac{\sqrt{nh}\left\|T_2-\mathbf{E}(Qf)\right\|_\infty}{\sqrt{\max\left(\ln(1/h),\ln\ln n\right)}}=k_2(c)<\infty. \quad (46)$$

where for sufficiently large n,

$$\left\|\left(\frac{T_2}{T_1}-F\right)-\frac{T_2-Qf}{f}+\frac{Qf}{f^2}(T_1-f)\right\|_\infty\le C_1\frac{\ln n}{k}. \quad (47)$$

Thus,

$$\sqrt{k}\left\|\left(\frac{T_2}{T_1}-F\right)-\frac{T_2-Qf}{f}+\frac{Qf}{f^2}(T_1-f)\right\|_\infty\underset{n\to\infty}{\overset{p}{\to}}0. \quad (48)$$

Further,

$$\mathbf{D}\left(\frac{T_2}{T_1}-Q\right)=\left(\frac{1}{f^2}\mathbf{D}(T_2-Qf)+\frac{(Qf)^2}{f^4}\mathbf{D}(T_1-f)-\right.$$
$$\left.-2\frac{Qf}{f^3}\mathbf{Cov}(T_1-f,T_2-Qf)\right)\left(1+O_p\left(\frac{\ln^2 n}{k^2}\right)\right)=$$
$$=\frac{1}{f^2}\mathbf{D}(T_2)+\frac{(Qf)^2}{f^4}\mathbf{D}(T_1)-$$
$$-2\frac{Qf}{f^3}\mathbf{Cov}(T_1,T_2)\left(1+O_p\left(\frac{\ln^2 n}{k^2}\right)\right) \quad (49)$$

as $n\to\infty$.

Consider the expectation $\mathbf{E}(T_1\cdot T_2)$.

We have

$$\mathbf{E}(T_1\cdot T_2)=\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n}L_\rho(U_i-x)\cdot\frac{1}{n}\sum_{i=1}^{n}W_iL_\rho(U_i-x)\right)=$$
$$=\frac{1}{n^2}\mathbf{E}\left(\sum_{i=j=1}^{n}W_i\left(L_\rho(U_i-x)\right)^2+\right.$$
$$\left.+\sum_{i\ne j}L_\rho(U_i-x)W_jL_\rho(U_j-x)\right). \quad (50)$$

By virtue of independent and identically distributed pairs, we conclude that

$$\mathbf{E}(T_1\cdot T_2)=\frac{1}{n}\mathbf{E}\left(W_1\left(L_\rho(U_1-x)\right)^2\right)+$$
$$+\frac{n-1}{n}\mathbf{E}\left(W_1L_\rho(U_1-x)\right)\mathbf{E}\left(L_\rho(U_2-x)\right)=$$
$$=\frac{1}{n}\int\mathbf{E}\left(I(U_1>X_1)L_\rho^2(U_1-x)\middle|U_1=u\right)f(u)\,du+$$
$$+\frac{n-1}{n}\int\mathbf{E}\left(I(X_1<U_1)L_\rho(U_1-x)\middle|U_1=u\right)f(u)\,du\times$$
$$\times\int L_\rho(u-x)f(u)\,du. \quad (51)$$

Making the replacement $z=(u-x)r^{-1}$, we have

$$\mathbf{E}(T_1\cdot T_2)=(nr^d)^{-1}\int L^2(u-x)Q(zr+x)\times$$
$$\times f(zr+x)\,dz+(r^{2d})^{-1}(1-n^{-1})\times$$
$$\times\left(\int L(z)Q(zr+x)f(zr+x)\,dz\right)\times$$
$$\times\left(\int L(z)f(zr+x)\,dz\right),$$

$$(nr^d)^{-1}\int L^2(u-x)Q(zr+x)f(zr+x)dz=$$

$$= k^{-1}Q(\boldsymbol{x})f^2(\boldsymbol{x})\|L\|^2 + o(k^{-1}). \qquad (52)$$

From the conditions on the kernel $L(\boldsymbol{x})$ and the conditions on the functions $Q(\boldsymbol{x}),\ f(\boldsymbol{x})$, we have

$$\int L(\boldsymbol{z})Q(\boldsymbol{z}r+\boldsymbol{x})f(\boldsymbol{z}r+\boldsymbol{x})d\boldsymbol{z} = Q(\boldsymbol{x})f(\boldsymbol{x}) + o(k^{-1}).(53)$$

Thus,

$$\mathbf{E}(T_1\cdot T_2) = Q(\boldsymbol{x})f^2(\boldsymbol{x})\|L\|^2 k^{-1} + ,$$

$$\mathbf{Cov}(T_1,T_2) = Q(\boldsymbol{x})f^2(\boldsymbol{x})\|L\|^2 k^{-1} + O(k^{-2}).$$

So, as $n\to\infty$,

$$\mathbf{D}(\hat{Q}_n(\boldsymbol{x}) - Q(\boldsymbol{x})) = Q(\boldsymbol{x})(1-Q(\boldsymbol{x}))\|L\|^2 k^{-1}(1+o(1)). \ (55)$$

Hence we conclude that

$$\sqrt{k}(\tilde{Q}_n(\boldsymbol{x}) - Q(\boldsymbol{x})) \underset{n\to\infty}{\overset{d}{\to}} \varsigma \in N(0, Q(\boldsymbol{x})(1-Q(\boldsymbol{x}))\|L\|^2). \ (56)$$

4.2. Indirect Data

Let $\Upsilon^{(n)} = \{(Y_i,U_i,W_i), i=1,..,n\}$ be a sequence of tree-dimensional i.i.d. random vectors with probability density function $q(y,u)$ of the pair (Y_1,U_1) respect to Lebegue measure and

$$p(y) = \int_{-\infty}^{\infty} f(y,u)\,du\,,\ \ f(u\,|\,y) = f(u,y)\,/\,p(y)\,.$$

Let

$$R(y) = \int_{-\infty}^{\infty} Q(u)\,f(u|y)\,du = \frac{m(y)}{p(y)},$$

$$m(y) = \int_{-\infty}^{\infty} Q(u)\,f(y,u)\,du,$$

$$\sigma^2(y) = \|L\|^2 \int_{-\infty}^{\infty} Q(u)(1-Q(u))\,f(u\,|\,y)\,du.$$

Consider the statistics

$$Q_n^* = Q_n^*(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{i/n - F_n(y)}{h}\right)W_n^{[i]} \qquad (57)$$

and

$$Q_n^{**} = Q_n^{**}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{i/n - F(y)}{h}\right)W_n^{[i]}, \qquad (58)$$

where $F_n(y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i < y)$ is the empirical distribution function of Y-observations y, $F(y) = \mathbf{P}(Y<y)$ and $W_n^{[i]}$ is *i*-th concomitants of order statistics $Y_n^{(i)}$.

Theorem 4. *Let $\{(W_i,Y_i), 1\le i\le n\}$ be a random sample of size n – that is, a sequence of the independent identically distributed two-dimensional random vectors, where $W_i = I(X_i < U_i)$. Suppose the following conditions satisfied: the conditional density $f(u\,|\,y)$ has boundared continuous derivatives to third order inclusive with respect to y at (x,y).*

Then

$$\frac{\sqrt{n}(Q_n^*(y) - R(y))}{\sigma(y)} \underset{n\to\infty}{\overset{d}{\to}} \xi \in N(0,1). \qquad (59)$$

Proof. Let's notice that Q_n^* and Q_n^{**} are asymptotically equivalent in mean square. To see this, write

$$\frac{1}{h}\left|K\left(\frac{i/n - F_n(y)}{n}\right) - K\left(\frac{i/n - F(y)}{n}\right)\right| \le$$
$$\le C_1 h^{-2}|F_n(y) - F(y)|. \qquad (60)$$

The law of the iterated logarithm imply that with probability one (7) converges to zero uniformly in i as $n\to\infty$. Hence, we may consider Q_n^{**} rather than Q_n^*. Let $v = F(y_0), y_0 = F^{-1}(v)$. We apply theorem 2.

Therefore

$$\mu = \mu_n = \frac{1}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{F(y)-v}{h}\right)\times$$
$$\times(I(X<Y)\,|\,Y=y, U=u)\,f(y,u)\,dydu =$$
$$= \frac{1}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{F(y)-v}{h}\right)F(u)\,f(u\,|\,y)p(y)\,dydu =$$
$$= \frac{1}{h}\int_{0}^{1}\int_{-\infty}^{\infty} K\left(\frac{t-v}{h}\right)F(u)\,f(u\,|\,F^{-1}(t))\,dtdu =$$
$$= \int_{-v/h}^{(1-v)/h}\int_{-\infty}^{\infty} K(z)F(u)\,f(u\,|\,F^{-1}(v+zh))\,dzdu.$$

For sufficiently large n we will have

$$\mu = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(z)F(u)\,f(u\,|\,F^{-1}(v+zh))\,dzdu. \quad (61)$$

If the function $g(u\,|\,F^{-1}(v+zh))$ is continuous function at each point z. Then we will have the following

$$\lim_{n\to\infty}\mu = \int_{-\infty}^{\infty} Q(u)\,f(u\,|\,F^{-1}(v+zh) =$$
$$= \int_{-\infty}^{\infty} Q(u)\,f(u\,|\,y_0) = R(y_0)\,. \qquad (62)$$

Furthermore,

$$h\cdot\sigma^2 = \frac{1}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}((F(y_1\wedge y_2) - F(y_1)F(y_2))\times$$
$$\times K\left(\frac{y_1-v}{h}\right)K\left(\frac{y_2-v}{h}\right)dF(y_1)dF(y_2) +$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2\left(\frac{F(y)-v}{h}\right)Q(u)(1-Q(u))\,f(y,u)\,dydu. \ (63)$$

For the first summand we have (as $n\to\infty$)

$$\frac{1}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}((F(y_1\wedge y_2) - F(y_1)F(y_2))\times$$

$$\times K\left(\frac{y_1-v}{h}\right)K\left(\frac{y_2-v}{h}\right)dF(y_1)dF(y_2)=o(h). \quad (64)$$

Reasoning as in the previous case we will receive

$$\int_{-\infty}^{\infty}\frac{1}{h^2}\int_{-\infty}^{\infty}K^2\left(\frac{F(y)-v}{h}\right)Q(u)(1-Q(u))\,f(y,u)\,dydu\sim$$

$$\sim\frac{\|K\|^2}{h}\int_{-\infty}^{\infty}Q(u)(1-Q(u))\,f(u\,|\,y_0)\,du. \quad (65)$$

From here

$$\frac{\sqrt{nh}(Q_n^*(y)-R(y))}{\sigma(y)}\xrightarrow[n\to h]{d}\xi\in N(0,1). \quad (66)$$

Let's notice that for finite-sample performance of the distribution function it is better to use *new Yang-type estimator* of the form

$$\tilde{Q}_n(x)=\sum_{i=1}^{n}K\left(\frac{i/n-F_n(x)}{h}\right)W_n^{[i]}\Bigg/\sum_{i=1}^{n}K\left(\frac{i/n-F_n(x)}{h}\right), \quad (67)$$

which an asymptotically equivalent to an estimator Q_n^*.

4.3. Indirect Quantiles

When $d=1$, i.e. when Y is real-valued, the function Q_n^* has an inverse or quantile function

$Q_n^{*-1}(\lambda)=\inf\{y\in R: Q_n^*(y)\ge\lambda\},\ \ 0<\lambda<1.$

This is scheduled for estimating the λ quantile $Q^{-1}(\lambda)$ of $Q(\cdot)$. In this subsection we derive the limit distribution of

$$\varsigma_n(\lambda)\equiv(nh)^{1/2}(Q_n^{*-1}(\lambda)-Q^{-1}(\lambda)),\ 0<\lambda<1 \quad (68)$$

fixed.

For such an λ, write $y_\lambda=Q^{-1}(\lambda)$.

Theorem 5. *Under the assumptions of the theorem 4, if $q(y_\lambda)=Q'(y)>0$ at $y=y_\lambda$ and $q(y,u)$ is continuous we have*

$$\varsigma_n(\lambda)\to N(0,\sigma_\lambda^2)\ \textit{in distribution}, \quad (69)$$

where

$$\sigma_\lambda^2=\lambda(1-\lambda)\|L\|^2/(q(y_\lambda))^2. \quad (70)$$

Proof. For proof see [22].

5 Reduction of a measurement error

Let the dose $\boldsymbol{U}$ is measured with an error, i.e. $\boldsymbol{Y}=\boldsymbol{U}+\boldsymbol{\varepsilon}$, where $\boldsymbol{U},\boldsymbol{\varepsilon}$ are independent random variables and $\boldsymbol{\varepsilon}\in\boldsymbol{R}^d$ has normal distribution with d-dimensional mean vector $\mathbf{0}$ and a known $d\times d$ covariance matrix $\boldsymbol{\Sigma}_0$, and the random vector $\boldsymbol{U}$ has unknown density $g(\boldsymbol{u})>0$. The regression curve of $\boldsymbol{U}$ with respect to $\boldsymbol{Y}$ it can be written in form

$$\boldsymbol{u}(\boldsymbol{x})=\mathbf{E}(\boldsymbol{U}\,|\,\boldsymbol{Y}=\boldsymbol{x})=\frac{\boldsymbol{r}(\boldsymbol{x})}{q(\boldsymbol{x})}, \quad (71)$$

where

$$\boldsymbol{r}(\boldsymbol{x})=\int\boldsymbol{u}g(\boldsymbol{u})\times$$

$$\times\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_0|^{1/2}}|\exp\left(-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{x})^T\boldsymbol{\Sigma}_0(\boldsymbol{u}-\boldsymbol{x})\right)d\boldsymbol{u}, \quad (72)$$

$$q(\boldsymbol{x})=\int g(\boldsymbol{u})\times$$

$$\times\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_0|^{1/2}}|\exp\left(-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{x})^T\boldsymbol{\Sigma}_0(\boldsymbol{u}-\boldsymbol{x})\right)d\boldsymbol{u}. \quad (73)$$

Differentiating $q(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ yields (see [23])

$$\nabla_q(\boldsymbol{x})=-\boldsymbol{\Sigma}_0^{-1}\boldsymbol{x}q(\boldsymbol{x})+\boldsymbol{\Sigma}_0^{-1}\boldsymbol{r}(\boldsymbol{x}), \quad (74)$$

where the symbol $\nabla_q(\boldsymbol{x})$ denote the $1\times d$ matrix of first-order partial derivatives of the transformation from $\boldsymbol{x}$ to $q(\boldsymbol{x})$.

Let the random vector $\boldsymbol{Y}$ has normal distribution with d-dimensional unknown mean vector $\boldsymbol{a}$ and a known $d\times d$ covariance matrix $\boldsymbol{\Sigma}$. Then

$$\boldsymbol{\Sigma}_0\frac{\nabla_q(\boldsymbol{x})}{q(\boldsymbol{x}}=-\boldsymbol{x}+\frac{\boldsymbol{r}(\boldsymbol{x})}{q(\boldsymbol{x}}=\nabla_{\ln q}(\boldsymbol{x})=\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{a}), \quad (75)$$

from where

$$\boldsymbol{\Sigma}_0\frac{\nabla_q(\boldsymbol{x})}{q(\boldsymbol{x}}+\boldsymbol{x}=(\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0)\boldsymbol{\Sigma}^{-1}\boldsymbol{x}-\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{a}. \quad (76)$$

Since $\boldsymbol{a}$ and $\boldsymbol{\Sigma}$ are unknown, we will estimate them on sample $\boldsymbol{y}_1,\boldsymbol{y}_2,...,\boldsymbol{y}_n$ with the help of the following the statistics

$$\hat{\boldsymbol{a}}=\overline{\boldsymbol{y}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i \quad (77)$$

and

$$\hat{\boldsymbol{\Sigma}}=\mathbf{S}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i-\overline{\boldsymbol{y}})(\boldsymbol{y}_i-\overline{\boldsymbol{y}})^T. \quad (78)$$

The regression estimation in this case will be equal

$$\hat{\boldsymbol{u}}_n(\boldsymbol{x})=(\mathbf{S}-\boldsymbol{\Sigma}_0)\mathbf{S}^{-1}\boldsymbol{x}+\boldsymbol{\Sigma}_0\mathbf{S}^{-1}\overline{\boldsymbol{y}}. \quad (79)$$

If instead of $\boldsymbol{x}$ we will substitute observable value $\boldsymbol{y}_i$, then the corrected value of a vector $\hat{\boldsymbol{u}}_i$ we calculate the corrected value of a vector $\hat{\boldsymbol{u}}_i$ using the formula

$$\hat{\boldsymbol{u}}_i=\hat{\boldsymbol{u}}_n(\boldsymbol{y}_i)=(\mathbf{S}-\boldsymbol{\Sigma}_0)\mathbf{S}^{-1}\boldsymbol{y}_i+\boldsymbol{\Sigma}_0\mathbf{S}^{-1}\overline{\boldsymbol{y}}. \quad (80)$$

6 Numerical properties

In this section we report the results of a simulation study illustrating the theoretical results and finite-sample behavior of the estimators of population features considered in Sections 2-5. We consider

For the error distribution, we consider the normal distribution $N(0,0.23^2)$. For each combination of the target and error distributions, we consider two different sample sizes, $n=100$ and

$n = 200$. In the simulation study for this section we choose the kernel

$$K(x) = \frac{15}{16}(1-x^2)^2 I(|x| \le 1) . \qquad (81)$$

We consider the case when the initial data does not include measurement error and also the case when a measurement error is superimposed on the initial data.

Case 1. Initial data without imposing a measurement error

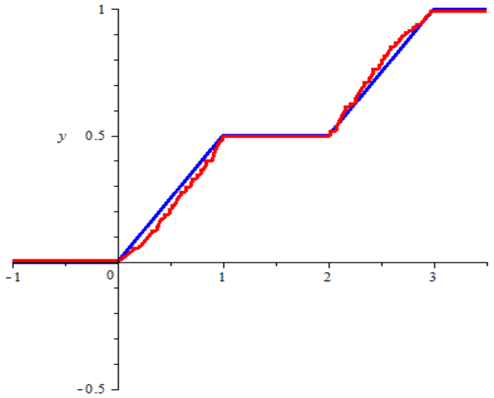

Fig. 1. Empirical distribution function and estimation of the initial data, $n = 100$

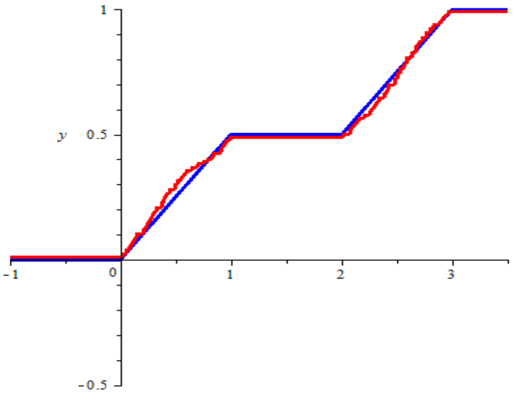

Fig. 2. Empirical distribution function and estimation of the initial data, $n = 200$.

Case 2. The initial data with the imposition of a measurement error

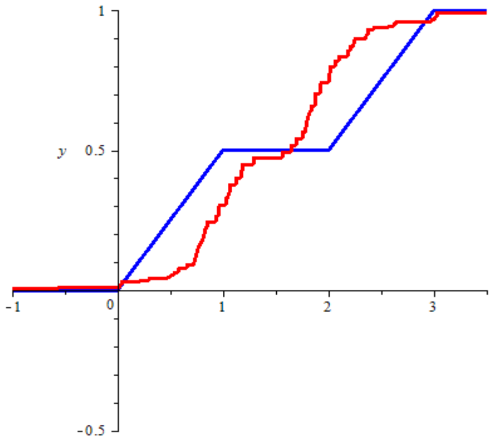

Fig. 3. Empirical distribution function and estimation according to the superimposed measurement error, $n = 100$

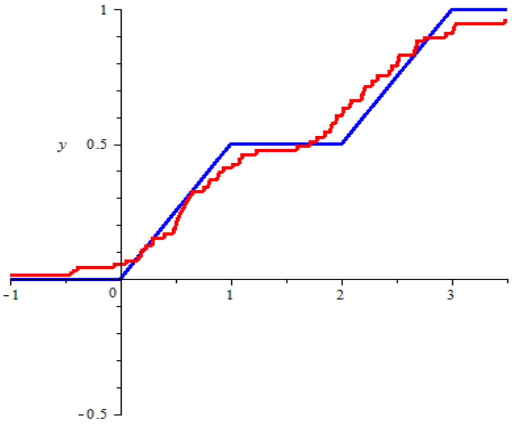

Fig. 4. Empirical distribution function and estimation according to the superimposed measurement error, $n = 200$

Case 3. The data overlay measurable error after conversion

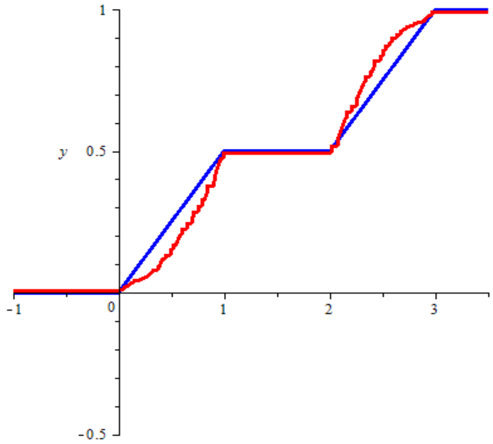

Fig. 5. Empirical distribution function and estimation according to the superimposed measurement error after conversion, $n = 100$

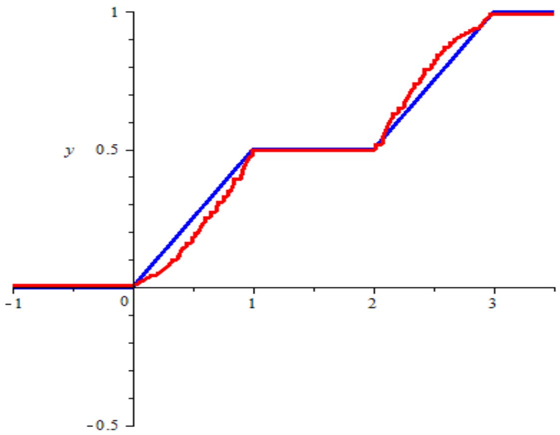

Fig. 6. Empirical distribution function and estimation according to the superimposed measurement error after conversion, $n = 200$

By construction, in the application package MatLab graphs easy to see that applying the error estimate of the distribution function does not match the empirical distribution function. But after converting the data with the superimposed measurement error estimation function distribution shows good results as seen from the Fig. 5 and Fig. 6.

7 Conclusion

In the case where the response is binary, one way to describe is to build a logistic regression, which assumes knowledge of the distribution close to her.

Another method is the nonparametric estimation method - kernel estimation.

It is usually assumed that the variables are measured without error. If the variables are measured with errors (convolution model), proceed as follows: build the kernel using the direct and inverse Fourier transform, which would compensate the measurement error. Our approach is based on the fact that we first reduce the measurement error, and then evaluate the new data measurement function.

The results of numerical simulations show the success of this approach.

References:

[1] D.J. Finney, *Probit-analysis*, Cambridge University Press, 1947.

[2] H. Dette, N. Neumeyer, K.F. Pilz, A note on nonparametric estimation of the effective dose in quantal bioassay, *Journal Amer. Statist. Assoc*., Vol.100, No.470, pp.503-510.

[3] M.S. Tikhov, Nonparametric estimation of effective doses at quantal response, *Ufimsk. Mat. Zh*, Vol. 5, Issue. 2, 2013, pp. 94-108.

[4] M.S. Tikhov, Asymptotical normality of *kNN*-estimates in dose-effect dependence, *Nizhny Novgorod State University Bulletin*. *Series: Mathematika*, No.1(4), 2006, pp. 129-137.

[5] D.J. Finney, *Statistical Method in Biological Assays*, Griffin, London, 1964.

[6] D.J. Finney, The choice of a response metameter in bioassay. *Biometrics*, Vol. 5, pp.261-272.

[7] C.C.Brown, Statistical Analysis of Dose-effect Relationships *In: Principles of Ecotoxicology*, Jonh Willey&Sons, NY, 1978, pp.115-148.

[8] M.Chu, K. P. Huynh, D.T. Jacho-Chavez, Functionals of order statistics and their multivariate concomitants with application to semiparametric estimation by nearest neighbours. *Sankhyā: The Indian Journal of Statistics*, Vol. 75-B, Part 2, 2013, pp.238-292.

[9] M.S. Tikhov, Statistical Estimation on the Basis of Interval-Censored Data. *Journal Math. Sciences*, Vol. 119, No.3, 2004, pp. 321-335.

[10] H. Midi, S. Rana, S.K. Sarkar, Binary Response Modeling and Validation of its Predictive Ability, *Wseas Transactions on Mathematics*, Vol.9, No.6, 2010, pp. 436-447.

[11] J. Yang, L. Fan, Weighted Generalized Kernel Discriminant Analysis Using Fuzzy Memberships, *Wseas Transactions on Mathematics*, Vol.10, No.10, 2011, pp. 346-357.

[12] J. Gao, L. Fan, Kernel-based Weighted Discriminant Analysis with QR Decomposition and Its Application to Face Recongnition, *Wseas Transactions on Mathematics*, Vol.10, No.10, 2011, pp. 358-367.

[13] S. Yang, Linear function of concomitants of order statistics with application to nonparametric estimation of a regression function, *Journal Amer. Statist. Assoc*, Vol. 76, 1981, pp. 658-662.

[14] M.S. Tikhov, *Statistical Estimation based on Interval Censored Data*. Param. and Semiparam. Models with Appl. to Rel., Surv. Analisys, and Qual. of Life: Springer-Verlag: Theor.& Meth, 2004, pp.209-215.

[15] M. Tikhov, M. Ivkin, Multivariate k-Nearest Neighbors Distribution Function Estimates in Dose-effect Relationship, *Recent Advances in Mathematical Methods in Applied Sciences: Proceedings of the 2014 International Conference on Mathematical Models and Methods in Applied Sciences*, 2014, pp.325-329.

[16] M.S. Tikhov, Statistical Estimation based on Interval Censored Data. – In: *Param. and Semiparam. Models with Appl. to Rel., Surv. Analisys, and Qual. of Life:* Springer-Verlag: Theor.& Meth., 2004, XLIV, pp. 209-215.

[17] P. Hall, S.N. Lahiri, Estimation of distribution, moments and quantiles in deconvolution problems, *Ann. Statist*., Vol. 36, 2008, pp.2110-2134.

[18] C.R. Rao, *Linear statistical inference and its applications.* New York: John Wiley & Sons, 1965.

[19] Y.P. Mack, M. Rosenblatt, Multivariate *k*-Nearest Neighbor Density Estimates. *Journal of Multivar. Analysis*, vol. 9, pp. 1-15, 1979.

[20] N.V. Smirnov, Limit distributions for the terms of a variational series. *Trudy Mat. Inst. Steklov., Acad. Sci. USSR*, Vol. 25, 1949, pp. 3–60.

[21] U. Einmahl, D.M. Mason, Uniform in Bandwidth Consistency of Kernel-type Function Estimators. *Ann. Statist*., vol. 33, no.3, pp.1380-1403, 2005.

[22] W. Stute, Conditional Empirical Process. *Ann. Statist*., Vol.14, 1986, pp. 638-647.

[23] J.R.Magnus, H.Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Third Edd., Jonson Wiley&Sons, NY, 2007, 450 p.