ᲡᲐᲥᲐᲠᲗᲕᲔᲚᲝᲡ ᲛᲔᲪᲜᲘᲔᲠᲔᲑᲐᲗᲐ ᲔᲠᲝᲕᲜᲣᲚᲘ ᲐᲙᲐᲓᲔᲛᲘᲘᲡ ᲛᲝᲐᲛᲑᲔ, ტ. 8, №2, 2014
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 8, no. 2, 2014

Mathematics

Integral Functionals of the Priestley-Chao Regression Function

Dimitri Arabidze

V. Komarov Public School of Physics and Mathematics, Tbilisi

(Presented by Academy Member Elizbar Nadaraya)

ABSTRACT. In the present paper the problem of statistical estimation of the nonlinear integral functional of a regression function is discussed. For the regression function and its derivatives well known Priestley-Chao estimator are taken. The problem is naturally considered in the Sobolev space. As an estimator for this function the plug-in estimatoris proposed. Theorems about consistency and asymptotically normality are proved. The order of the convergence is determined. The general methodology is used for some special cases. The estimation problem of Fisher's information and Shannon entropy for Priestley-Chao's regression function is solved. *© 2014 Bull. Georg. Natl. Acad. Sci.*

Key words: *Priestley-Chao estimation, regression function, integral functional.*

In the present paper we investigate the integral functional of a regression function and its derivatives. In our investigation we use the Priestley-Chao Regression Function introduced and studied in [1-3].

The study of functionals of a probability distribution density or of a regression function and its derivatives is an interesting task and attracts an active interest of the part researchers [2-8]. Detailed studies of functionals of a probability distribution density function and its derivatives are presented [4-7]. Investigations of functionals of a regression function and its derivatives are more modest [2,3].

Let $a(t)$ denote the regression function, then we may consider, say, the particular cases:

$$I_1(a)=\int_{-\infty}^{\infty} a^2(t)dt, \qquad I_2(a)=\int_{-\infty}^{\infty} \frac{\left(a'(t)\right)^2}{a(t)}dt,$$

$$I_3(a)=\int_{-\infty}^{\infty} \left(a(t)\right)^s dt, \qquad I_4(a)=\int_{-\infty}^{\infty} a(t)\log a(t)dt,$$

Related problems were studied in the above-mentioned works [2,3]. Our approach in this paper is based on the derivation of a representation theorem which we further use to obtain the results connected with asymptotic properties, in particular with consistency and the central limit theorem. The statement of the problems and the discussion were inspired by [4].

Let us consider a regression model of the form:

$$Y(t) = a(t) + \varepsilon(t) \tag{1}$$

where $t \in [0,1]$, $\varepsilon(\cdot)$ is noise with $E\varepsilon(t) = 0$, $E\varepsilon^2(t) = \sigma^2 < \infty$, $Y(t)$ is an observed random function, and $a(t)$ is an unknown regression function. Suppose that we have n numbers:

$$0 \le t_1 \le t_2 \le ... \le t_n \le 1,$$

where each t_k, $k = 1,2,...,n$ depends on n and $\max_i |t_i - t_{i-1}| = O\left(\frac{1}{n}\right)$. We have n observations:

$$Y(t_1), Y(t_2), ..., Y(t_n).$$

The estimator of the unknown function $a(t)$ was introduced by Priestley M. B. and Chao M. T. [1] and defined by the expression:

$$\hat{a}_n(t) = \sum_{i=1}^{n} W\left(\frac{t - t_i}{h_n}\right) \cdot \frac{t_i - t_{i-1}}{h_n} \cdot Y(t_i), \tag{2}$$

where $\{h_n,\ n = 1,2,...\}$ is a sequence of positive numbers monotonically tending to zero. $W(t)$ is the function with probability density properties. In [1] the estimator of the k-th derivative of the regression function $a^{(k)}(t)$ is introduced as formula:

$$\hat{a}_n^{(k)}(t) = \frac{1}{h_n^k} \cdot \sum_{i=1}^{n} W^{(k)}\left(\frac{t - t_i}{h_n}\right) \cdot \frac{t_i - t_{i-1}}{h_n} \cdot Y(t_i), \tag{3}$$

for all $k = 0,1,2,...,m.$ It was assumed that $\hat{a}_n^{(0)}(t) \doteq \hat{a}_n(t)$.

Let $\varphi : \mathbb{R}^{m+2} \to \mathbb{R}$ be a continuous bounded function. Consider an integral functional of the form:

$$I(a) = \int_{-\infty}^{\infty} \varphi\left(t, a(t), a'(t), ..., a^{(m)}(t)\right) dt. \tag{4}$$

We have the selection (t_i, Y_i), $i = 1,2,...,n.$ This means that

$$Y_i = Y(t_i) = a(t_i) + \varepsilon(t_i). \tag{5}$$

To estimate $I(a)$ we use the plug-in estimator, i.e. consider the functional:

$$I(\hat{a}_n) = \int_{-\infty}^{\infty} \varphi\left(t, \hat{a}_n(t), \hat{a}_n'(t), ..., \hat{a}_n^{(m)}(t)\right) dt.$$

Representation Theorem

Our consideration is based on a representation theorem which will lead to the results we are interested in. Let us list the conditions, which the considered variables are supposed to satisfy.

Conditions on a:

(a1) The function a=a(t) is defined and continuous on [0, 1] and takes its values in the interval [-k,k];

(a2) a=a(t) has continuous derivatives up to order m inclusive;

(a3) For any $i = 0,1,2,...,m,$ $a^{(i)}(t)$ takes its values in [-k,k] and $a^{(i)}(\cdot) \in L_1([0,1])$.

Conditions on ε_k:

(ε1) Random values ε_k, k=1,2,... are independent, bounded and equally distributed;

(ε2) $E\varepsilon_k = 0,\ E\varepsilon_k^2 = \sigma^2 < \infty;$

For brevity, we will use notation for $\varphi = \varphi(x, x_0, ..., x_m) \in C_b^2(R^{m+2})$ function:

$$\frac{\partial\varphi}{\partial x_i} = \varphi_{(i)},\ i = 0,1,...,m \text{ and } \frac{\partial^2\varphi}{\partial x_i \partial x_j} = \varphi_{(ij)},\ i, j = 0,1,...,m.$$

Conditions on φ:

(φ1) The function $\varphi : R^{m+2} \to R$ is continuous, bounded, integrable and has bounded continuous derivatives up to second order, inclusive, in some convex domain A, which contains the domain $R \times [-k,k]^{m+1}$;

(φ2) All first and second derivatives of the function φ are uniformly bounded in the domain A by a constant $C_\varphi > 0$.

By this conditions for the function j we have for all $i, j = 0,1,...,m$:

$$\sup\{|\varphi_{(i,j)}|(s, s_0, s_1, ..., s_m) : (s, s_0, s_1, ..., s_m) \in A\} \le C_\varphi. \tag{6}$$

Conditions on W:

(W1) $\int_{-\infty}^{\infty} W(t)dt = 1;$

(W2) Function $W(t)$ has the compact support $[-\tau, \tau]$ and $W(-\tau) = W(\tau) = 0;$

(W3) $W(t)$ has continuous derivatives up to order $m \ge 1;$

(W4) There exists a constant $C_W > 0$, for which $\sup_{t\in R} |W^{(i)}(t)| \le C_W < \infty,\ i = 0,1,...,m;$

(W5) For any $i = 0,1,...,m,\ W^{(i)} \in L_1([-\tau, \tau]).$

Conditions on h_n:

$$(h_n 1)\ \frac{\sqrt{\max(|\log h_n|; \log\log n)}}{\sqrt{n} h_n^{0,5+m}} \to 0, \text{ as } n \to \infty.$$

Denote by $a_n(t)$ mathematical expectation $\hat{a}_n(t)$:

$$a_n(t) = E\hat{a}_n(t) = E\left(\sum_{i=1}^n W\left(\frac{t - t_i}{h_n}\right) \cdot \frac{t_i - t_{i-1}}{h_n} \cdot Y(t_i)\right) = \sum_{i=1}^n W\left(\frac{t - t_i}{h_n}\right) \cdot \frac{t_i - t_{i-1}}{h_n} \cdot a(t_i).$$

Then we obtain

$$a_n^k(t) = E\hat{a}_n^k(t) = \frac{1}{h_n^k} \cdot \sum_{i=1}^n W^{(k)}\left(\frac{t - t_i}{h_n}\right) \cdot \frac{t_i - t_{i-1}}{h_n} \cdot a(t_i).$$

Let us show that there also exist expressions $I(a)$, $I(a_n)$ and $I(\hat{a}_n)$ and they are finite.

Using the Taylor formula for any point $(s, s_0, s_1, ..., s_m) \in A$ and some $\tilde{S}_i \in A$ we can write

$$|\varphi|(s, s_0, s_1, ..., s_m) = \left|\sum_{t=0}^m \varphi_{(i)}(s, 0, 0, ..., 0)s_i + \frac{1}{2}\sum_{t=0}^m \varphi_{(i,j)}(s, \tilde{s}_0, \tilde{s}_1, ..., \tilde{s}_m)s_i s_j\right|.$$

Accordingly, there exists a constant C such that

$$|\varphi|(s,s_0,s_1,...,s_m)\le C\left(\sum_{t=0}^{m}|s_i|+\sum_{t=0}^{m}|s_i|^2\right).$$

Hence it follows that for any bounded measurable functions $f_0(t), f_1(t),..., f_m(t)$ from $L_1(R)$ we have

$$\int_{-\infty}^{\infty}|\varphi|(t,f_0(t),f_1(t),...,f_m(t))dt<\infty. \tag{7}$$

And therefore $I(a)$ exists.

The conditions which are imposed on the function W ensure boundness and membership in $L_1(R)$, then condition ($W4$) and (6)-(7) imply the finiteness of both variables $I(a_n)$ and $I(\hat{a}_n)$, for any $n\in N$.

By the Taylor formula we can write

$$I(\hat{a}_n)-I(a_n)=S_n(h_n)+R_n, \tag{8}$$

where, for any $h_n>0$, $S_n(h_n)$ is the sum of independent random variables:

$$S_n(h_n)=\sum_{i=0}^{m}\int_0^1\varphi_{(i)}\left(t,a_n(t),a_n'(t),...,a_n^m(t)\right)\left(\hat{a}_n^{(i)}(t)-a_n^{(i)}(t)\right)dt. \tag{9}$$

A remainder R_n has the form:

$$R_n=\frac{1}{2}\cdot\sum_{i,j=0}^{m}\int_0^1\varphi_{(ij)}\left(\tilde{b}_m(t)\right)\left(\hat{a}_n^{(i)}(t)-a_n^{(i)}(t)\right)\left(\hat{a}_n^{(j)}(t)-a_n^{(j)}(t)\right)dt. \tag{10}$$

Where $\tilde{b}_m(t)$ is a point on the straight line connecting the points

$$\left(t,\hat{a}_n(t),\hat{a}_n'(t),...,\hat{a}_n^{(m)}(t)\right)\text{and}\left(t,a_n(t),a_n'(t),...,a_n^{(m)}(t)\right).$$

Let us estimate the remainder R_n. Applying the standard procedure, from (7) and (10) we obtain:

$$R_n\le C_\varphi\cdot\int_0^1\sum_{i=0}^{m}\left(\hat{a}_n^{(i)}(t)-a_n^{(i)}(t)\right)^2dt. \tag{11}$$

Let W_m^2 denote the Sobolev space of functions having a square-integrable continuous and bounded second derivative with the norm $\|g\|_m=\sqrt{\sum_{i=0}^{m}\int_0^1\left|g^{(i)}(t)\right|^2dt}$ and the scalar product

$$\langle g_1,g_2\rangle_m=\sqrt{\sum_{i=0}^{m}\int_0^1\left(g_1^{(i)}(t)\cdot g_2^{(i)}(t)\right)dt}.$$

Denote $r_n(m)=\|\hat{a}_n-a_n\|_m^2$. Then we can write

$$|R_n|\le C_\varphi r_n(m). \tag{12}$$

Assume

$$U_k=U_k(t)=W\left(\frac{t-t_k}{h_n}\right)\cdot\frac{t_k-t_{k-1}}{h_n}\cdot\left[Y(t_k)-a(t_k)\right],\quad k=1,2,...,n,$$

where $a(t_k)=EY(t_k)$. Then

$$\sum_{k=1}^{n}U_k=\sum_{k=1}^{n}W\left(\frac{t-t_k}{h_n}\right)\cdot\frac{t_k-t_{k-1}}{h_n}\cdot\left[Y(t_k)-a(t_k)\right]=\hat{a}_n(t)-a_n(t).$$

Therefore,

$$r_n(m)=\left\|\sum_{k=1}^{n}U_k\right\|_m^2. \tag{13}$$

Let us estimate the norm of one of the summands U_k in (13) for each k=1,2,...,n. We obtain

$$\|U_k\|_m=\left(\sum_{i=0}^{m}\int_0^1\left|U_k^{(i)}(t)\right|^2dt\right)^{\frac{1}{2}}=\left(\sum_{i=0}^{m}\int_0^1\left|W\left(\frac{t-t_k}{h_n}\right)\cdot\left(\frac{t_k-t_{k-1}}{h_n}\cdot\left[Y(t_k)-a(t_k)\right]\right)^{(i)}\right|^2dt\right)^{\frac{1}{2}}$$

$$=\left(\sum_{i=0}^{m}\int_0^1\left|\frac{1}{h_n^i}\cdot W^{(i)}\left(\frac{t-t_k}{h_n}\right)\cdot\frac{t_k-t_{k-1}}{h_n}\cdot\left[Y(t_k)-a(t_k)\right]\right|^2dt\right)^{\frac{1}{2}}$$

$$=\left(t_k-t_{k-1}\right)\left[Y(t_k)-a(t_k)\right]\left(\sum_{i=0}^{m}\int_0^1\left|\frac{1}{h_n^{i+1}}\cdot W^{(i)}\left(\frac{t-t_k}{h_n}\right)\right|^2dt\right)^{\frac{1}{2}}\le\frac{|\varepsilon_k|C_W}{n}\cdot\left(\sum_{i=0}^{m}\int_0^1\left|\frac{1}{h_n^{i+1}}\right|^2dt\right)^{\frac{1}{2}}$$

$$=\frac{|\varepsilon_k|C_W}{n}\cdot\left(\frac{1-h_n^{2m+2}}{h_n^{2m+2}(1-h_n^2)}\right)^{\frac{1}{2}}\le L\cdot\frac{1}{nh_n^{m+1}}:=M_m\sim O\left(\frac{1}{nh_n^{m+1}}\right) \tag{14}$$

for sufficiently large L>0.

To estimate $r_n(m)$ we use the McDiarmid's inequality, which we give here for convenience (for details see [9]).

McDiarmid's Inequality: Let $H(t_1,t_2,...,t_k)$ be a real function such that for each k=1,2,...,n and some c_i, the supremum in $t_1,t_2,...,t_k,t$ of the difference

$$\left|H(t_1,t_2,...,t_{i-1},t_i,t_{i+1},...,t_k)-H(t_1,t_2,...,t_{i-1},t,t_{i+1},...,t_k)\right|\le c_i.$$

If $X_1,X_2,...,X_k$ are independent random variables taking values in the domain of the function $H(t_1,t_2,...,t_k)$ then for every $\varepsilon>0$

$$P\{|H(X_1,X_2,...,X_k)-EH(X_1,X_2,...,X_k)|>\varepsilon\}\le 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^{k}c_i^2}\right).$$

Let us apply McDiarmid's inequality for the function

$$H(U_1,U_2,...,U_n)=\left\|\sum_{k=1}^{n}U_k\right\|_m.$$

We have:

$$\left|H(U_1,U_2,...,U_{i-1},U_x,U_{i+1},...,U_n)-H(U_1,U_2,...,U_{i-1},U_y,U_{i+1},...,U_n)\right|=$$

$$=\left|\left\|\sum_{k=1,k\neq i}^{n}U_k+U_x\right\|_m-\left\|\sum_{k=1,k\neq i}^{n}U_k+U_y\right\|_m\right|\leq\|U_x\|_m+\|U_y\|_m\leq 2M_m.$$

And as c_k we take $c_k\equiv 2M_m,\ k=1,...,n,$ from (14), for any $\delta>0$ we obtain:

$$P=\left\{\left|\left\|\sum_{k=1}^{n}U_k\right\|_m-E\left\|\sum_{k=1}^{n}U_k\right\|_m\right|>\delta\right\}\leq 2\exp\left(-\frac{2\delta^2}{\sum_{k=1}^{n}4M_m^2}\right)=2\exp\left(-\frac{2\delta^2}{n\cdot 4M_m^2}\right)$$

$$=2\exp\left(-\frac{2\delta^2n^2h_n^{2m+2}}{n\cdot 4L^2}\right)=2\exp\left(-\frac{2\delta^2n^2h_n^{2m+2}}{2L^2}\right).$$

We substitute here $\delta=\dfrac{2L\sqrt{\log n}}{\sqrt{n}h_n^{m+1}}$

and we have:

$$P=\left\{\left|\left\|\sum_{k=1}^{n}U_k\right\|_m-E\left\|\sum_{k=1}^{n}U_k\right\|_m\right|>\delta\right\}\leq 2\exp\left(-\frac{4L^2\log nh_n^{2m+2}}{nh_n^{2m+2}\cdot 2L^2}\right)=2\exp\left(-2\log n\right)=\frac{2}{n^2}.$$

by the Borelli-Cantelli lemma, we write

$$\left\|\sum_{k=1}^{n}U_k\right\|_m=E\left\|\sum_{k=1}^{n}U_k\right\|_m+O\left(\frac{\sqrt{\log n}}{\sqrt{n}h_n^{m+1}}\right).\tag{15}$$

Using the Jensen's inequality

$$\left(E\left\|\sum_{k=1}^{n}U_k\right\|_m\right)^2\leq E\left\|\sum_{k=1}^{n}U_k\right\|_m^2=E\sum_{k=1}^{n}\sum_{i=0}^{m}\int_0^1\left|\frac{1}{h_n^i}\cdot W^{(i)}\left(\frac{t-t_k}{h_n}\right)\cdot\frac{t_k-t_{k-1}}{h_n}\cdot\left[Y(t_k)-a(t_k)\right]\right|^2dt\leq$$

$$\leq\frac{C_W^2}{n}\sum_{k=1}^{n}\sum_{i=0}^{m}\int_0^1\frac{1}{h_n^{2i+2}}\cdot E\left[Y(t_k)-a(t_k)\right]^2dt\leq\frac{C_W^2\delta^2}{n}\cdot\frac{1-h_n^{2m+2}}{h_n^{2m+2}(1-h_n^2)}\leq K\cdot\frac{1}{nh_n^{2m+2}}\tag{16}$$

from (12), (13), (15) and (16) we conclude that $R_n=O\left(\dfrac{\log n}{nh_n^{2m+2}}\right).$

Therefore the following statement is true.

Theorem 1.*Assume that conditions (a1)-(a3), (ε1)-(ε3), (φ1)-(φ3), (W1)-(W5) and (h1)are fulfilled.Then representation (8) is true and the remainder with probability 1 has the order*

$$R_n=O\left(\frac{\log n}{nh_n^{2m+2}}\right).\tag{17}$$

Consistency

In this section of the paper we use Theorem 1 to prove that the estimator $I(\hat{a}_n)$ is consistent.

Theorem 2. *Let the conditions of Theorem 1 be fulfilled. If the positive sequence* $(h_n)_{n=1}^{\infty}$, $0 < h_n < 1$ *is chosen so that*

$$\frac{\log n}{nh_n^{2m+2}} \to 0, \quad \text{as } n \to \infty \tag{18}$$

then with probability 1 we have

$$I(\hat{a}_n) \to I(a_n). \tag{19}$$

Proof. By Theorem 1 and formula (8)

$$I(\hat{a}_n) - I(a_n) = S_n(h_n) + R_n, \tag{20}$$

where $R_n = o(1)$ and

$$S_n(h_n) = \sum_{i=0}^{m} \int_0^1 \varphi(i)\left(t, a_n(t)a_n'(t), ..., a_n^{(m)}(t)\right)\left(\hat{a}_n^{(i)}(t) - a_n^{(i)}(t)\right)dt.$$

By condition $(a1)$:

$$\left\{\left(t, a_n(t)a_n'(t), ..., a_n^{(m)}(t)\right) : t \in [0,1]\right\} \subset [0,1] \times [-k,k]^{m+1}.$$

This and condition $(\varphi 2)$ imply that there exists a constant $C_\varphi > 0$, such that

$$\sup\left\{|\varphi_i|(t, t_0, t_1, ..., t_m) : (t, t_0, t_1, ..., t_m) \in [0,1] \times [-k,k]^{m+1}\right\} \le C_\varphi.$$

We can write:

$$ES_n(h_n) = 0.$$

$$DS_n(h_n) = ES_n^2(h_n) \le C_\varphi^2 \sum_{i=0}^{m} \int_0^1 E\left[\hat{a}_n^{(i)}(t) - a_n^{(i)}(t)\right]dt$$

$$\le C_\varphi^2 \sum_{i=0}^{m} \int_0^1 \left|\sum_{k=0}^{n} \frac{1}{h_n^i} \cdot W^{(i)}\left(\frac{t - t_k}{h_n}\right) \cdot \frac{t_k - t_{k-1}}{h_n} \cdot E\left[Y(t_k) - a(t_k)\right]\right|^2 dt$$

$$\le C_\varphi^2 C_W^2 \delta^2 \cdot \sum_{k=0}^{m} \left(t_k - t_{k-1}\right)^2 \cdot \left(\frac{1}{h_n^{i+1}}\right)^2 \sim C\frac{1}{nh_n^{2m+2}} \to 0 \tag{21}$$

because $\frac{\log n}{nh_n^{2m+2}} \to 0$ and do $S_n(h_n) \to 0$ as $n \to \infty$.

We can write

$$Ea_n^{(k)}(t) = \int_{-\tau}^{\tau} W(u)a^{(k)}(t)(t - uh_u)du + O\left(\frac{1}{nh_n^k}\right). \tag{22}$$

Hence we make the following conclusions:

i) for conclusion (17), $\frac{1}{nh_n^k}$ tends to zero for any k=0,1,...,m;

ii) $Ea_n^{(k)}(t) \to a^{(k)}(t)$ as $n \to \infty$.

Summarizing the above discussion, we ascertain that if $n \to \infty$ then

$$I(a_n) = \int_0^1 \varphi\left(t, a_n(t), a_n'(t), \ldots, a_n'^{(m)}(t)\right) dt \to \int_0^1 \varphi\left(t, a(t), a'(t), \ldots, a'^{(m)}(t)\right) dt = I(a).$$

Since $I(\hat{a}_n) - I(a_n) = o(1),$ we conclude that $I(\hat{a}_n) - I(a_n) \to 0$ a.e.The theorem is proved.

Central Limit Theorem

Using our representation theorem we can obtain the limit distribution property for the integral functional

$$I(\hat{a}_n) = \int_0^1 \varphi\left(t, \hat{a}_n(t), \hat{a}_n'(t), \ldots, \hat{a}_n^m(t)\right) dt.$$

Consider the difference

$$I(\hat{a}_n) - I(a_n) = S_n(h_n) + R_n, \tag{8}$$

where for any $h_n > 0,\ S_n(h_n)$ is the sum of independent random variables

$$S_n(h_n) = \sum_{i=0}^{m} \int_0^1 \varphi_{(i)}\left(t, a_n(t), a_n'(t), \ldots, a_n'^{(m)}(t)\right)\left(\hat{a}_n^{(i)}(t) - a_n^{(i)}(t)\right) dt. \tag{9}$$

R_n is a remainder having the form:

$$R_n = \frac{1}{2} \cdot \sum_{i,j=0}^{m} \int_0^1 \varphi_{(ij)}\left(\tilde{b}_m(t)\right)\left(\hat{a}_n^{(i)}(t) - a_n^{(i)}(t)\right)\left(\hat{a}_n^{(j)}(t) - a_n^{(j)}(t)\right) dt. \tag{10}$$

Clearly,

$$ES_n(h_n) = 0 \text{ and } ER_n \to 0 \text{ as } n \to \infty \tag{23}$$

Moreover

$$E\left(S_n(h_n)\right)^2 = \sigma^2 \left(\sum_{i=0}^{m} \int_0^1 \varphi_{(i)}\left(t, a_n(t), a_n'(t), \ldots, a_n'^{(m)}(t)\right) dt \right)^2. \tag{24}$$

and $\mathrm{Var} R_n \to 0$ as $n \to \infty$.

Using appropriate conditions, we have to prove that the variable $\sqrt{n}(I(\hat{a}_n) - I(a_n))$ is asymptotically normal and calculate the limiting variance. For this, according to the theorem and formulas (8), (23) and (24), we have to show the asymptotic normality of the variable $\sqrt{n} S_n(h_n)$. As follows from (10), in this case it suffices to study this property for the variables:

$$d_k = Y(t_k) \cdot \sum_{i=0}^{m} \frac{1}{h_n^i} \int_0^1 W^{(i)}\left(\frac{t - t_k}{h_n}\right) \cdot \frac{t_k - t_{k-1}}{h_n} \cdot \varphi_{(i)}\left(t, a_n(t), a_n'(t), \ldots, a_n^{(m)}(t)\right) dt \tag{25}$$

It can be easily verified that

$$Ed_k = a(t_k) \cdot \sum_{i=0}^{m} \frac{1}{h_n^i} \int_0^1 W^{(i)}\left(\frac{t - t_k}{h_n}\right) \cdot \frac{t_k - t_{k-1}}{h_n} \cdot \varphi_{(i)}\left(t, a_n(t), a_n'(t), \ldots, a_n^{(m)}(t)\right) dt. \tag{26}$$

Thus we consider the sequence of independent random variables:

$$f_n(k) = \alpha(n,k)\left(Y(t_k) - a(t_k)\right) = \alpha(n,k)\varepsilon_k,$$

where

$$a(n,k) = \sum_{i=0}^{m} \frac{1}{h_n^i} \cdot \int_0^1 W^{(i)}\left(\frac{t-t_k}{h_n}\right) \cdot \frac{t_k - t_{k-1}}{h_n} \cdot \varphi_{(i)}\left(t, a_n(t), a_n'(t), \ldots, a_n^{(m)}(t)\right) dt.$$

Let consider the sum $S_n(h_n) = \sum_{k=1}^{n} \alpha(n,k)\varepsilon_k$.

Let $F_{k,n}$ be the probability distribution function of a random variable $\alpha(n,k)\varepsilon_k$ and F_ε be the distribution function of a random variable - ε_k. The Linderberg's condition is written in the form $\forall \delta > 0, \lim_{n\to\infty} L_n(\delta) = 0$ where

$$L_n(\delta) = \frac{\sum_{j=1}^{n} \int x^2 J\left(|x| \geq \delta\sigma\left(\sum_{k=1}^{n} \alpha^2(n,k)\right)^{1/2}\right) dF_{k,n}(x)}{\sigma^2 \sum_{k=1}^{n} \alpha^2(n,k)},$$

here *J*(*A*) is the indicator function of the set *A*. It is easy to see that

$$L_n(\delta) \leq \frac{1}{\sigma^2} \max_{1\leq j\leq n} \int x^2 J\left(|x| \geq \delta\sigma v(n,j)\right) dF_\varepsilon(x)$$

where

$$v(n,j) = \frac{|\alpha(n,j)|}{\left(\sum_{j=1}^{n} \alpha^2(n,j)\right)^{1/2}}.$$

It remains to show that $\max_{1\leq j\leq n} v(n,j) \to 0$ as $n \to \infty$. But since

$$\max_{1\leq j\leq n} |\alpha(n,j)| = O\left(\frac{1}{nh_n^{m+1}}\right),$$

we have

$$\max_{1\leq j\leq n} v(n,j) = O\left(\frac{1}{\sqrt{n}}\right).$$

Thus the Linderberg's condition is fulfilled and we can conclude that the theorem is valid.

Theorem 3: *Let the conditions of Theorem 1 be fulfilled. Then if* $h_n \to 0$ *and* $\sqrt{n}h_n^{m+1} \to \infty$ *as* $n \to \infty$, *we have* $\sqrt{n}\left(I(\hat{a}_n) - I(a_n)\right) \to_d N(0, r^2)$, *where*

$$r^2 = \sigma^2 \cdot \sum_{i=0}^{m} \left(\int_0^1 \varphi_{(i)}\left(t, a_n(t), a_n'(t), \ldots, a_n^{(m)}(t)\right) dt \right)^2.$$

Applications

Let us consider the integral functional $I_1(a)=\int_0^1 a^2(t)dt.$

Then $\varphi(t,x_0,x_1,...,x_m)=x_0^2$ for $x_0\in[-b,b]\supset[-k,k],\ b>0.$ Thus $r^2=4\sigma^2\left(\int_0^1 a(t)dt\right)^2.$

And, using the conditions $h_n\to 0$ and $\sqrt{n}h_n^{m+1}\to\infty$ as $n\to\infty$, we have the convergence

$$\sqrt{n}\left(I(\hat{a}_n)-I(a_n)\right)\to_d N(0,r^2).$$

For the functional $I_2(a)=\int_0^1 \frac{(a'(t))^2}{a(t)}dt$ we obtain $\varphi(t,x_0,x_1,...,x_m)=\frac{x_1^2}{x_0}.$

Then, assuming that $t\in[0,1]\Rightarrow a(t)\in[a,b],\ \ b>a>0,$ we have

$$r^2=\sigma^2\cdot\left(-\int_0^1\left(\frac{(a'(t))^2}{(a(t))^2}-\frac{2a'(t)}{a(t)}\right)dt\right)^2=\sigma^2\cdot\left(\frac{a'(1)}{a(1)}\log a(1)-\frac{a'(0)}{a(0)}\log a(0)-\frac{1}{2}\cdot\frac{(a'(1))^2}{(a(1))^2}+\frac{1}{2}\cdot\frac{(a'(0))^2}{(a(0))^2}\right)^2.$$

For $h_n\to 0$ and $\sqrt{n}h_n^2\to\infty$ as $n\to\infty$, we have the convergence

$$\sqrt{n}\left(I(\hat{a}_n)-I(a_n)\right)\to_d N(0,r^2).$$

Let us consider the functional $I_3(a)=\int_{-\infty}^{\infty}(a(t))^s dt,\ s>1.$ Then $\varphi(t,x_0,x_1,...,x_m)=x_0^S,$ for $x_0\in[-b,b]\supset[-k,k],\ \ b>0.$ Therefore

$$r^2=s^2\sigma^2\cdot\left(\int_0^1 a^{s-1}(t)dt\right)^2.$$

And for the condition $h_n\to 0$ and $\sqrt{n}h_n\to\infty$ as $n\to\infty$,we have the convergence

$$\sqrt{n}\left(I(\hat{a}_n)-I(a_n)\right)\to_d N(0,r^2).$$

Let us now take the functional $I_4(a)=\int_{-\infty}^{\infty}a(t)\log a(t)dt.$ Then for some sufficiently large $b\geq K>0,$ if $0<x_0\leq b$ we have $\varphi(t,x_0,x_1,...,x_m)=\varphi(x_0)=x_0\log x_0.$ Let us extend the definition of the function φ by defining $\varphi(x)=0$ for $-b\leq x\leq b.$ Assume that $t\in[0,1]\Rightarrow a(t)\in[a,b],\ \ b>a>0$ and $b\geq b.$ Then

$$r^2=\sigma^2\cdot\left(\int_0^1 a(t)\left(1+\log a(t)\right)dt\right)^2.$$

And for the condition $h_n\to 0$ and $\sqrt{n}h_n\to\infty$ as $n\to\infty$, we have the convergence

$$\sqrt{n}\left(I(\hat{a}_n)-I(a_n)\right)\to_d N(0,r^2).$$

Iterated Logarithm Law

Applying the well-known iterated logarithm law from Kuelb's (paper [10]), we ascertain that the following statement is true.

Theorem 4. If *the sequence* h_n *is chosen so that*

$$R_n = o\left(\sqrt{\frac{\log\log n}{n}}\right),$$

then

$$\limsup_{n\to\infty} \pm \frac{\sqrt{n}\left(I(\hat{a}_n) - I(a_n)\right)}{\sqrt{2\log\log n}} = r.$$

Indeed, it can be easily verified that

$$\limsup_{n\to\infty} \pm \frac{\sqrt{n}\left(I(\hat{a}_n) - I(a_n)\right)}{\sqrt{2\log\log n}} = \limsup_{n\to\infty} \pm \frac{\sqrt{n}\left(a(n,k)Y(t_k) - a(n,k)a(t_k)\right)}{\sqrt{2\log\log n}} = r.$$

მათემატიკა

პრისტლი-ჩაოს რეგრესიის ფუნქციის ინტეგრალური ფუნქციონალების შესახებ

დ. არაბიძე

ვ. კომაროვის ფიზიკა-მათემატიკის საჯარო სკოლა, თბილისი,

ნაშრომში განხილულია რეგრესიის ფუნქციის ინტეგრალური სახის არაწრფივი ინტეგრალური ფუნქციონალების სტატისტიკური შეფასების პრობლემა. თვით რეგრესიის ფუნქციისათვის და მისი წარმოებულებისათვის აღებულია პრისტლი-ჩაოს ცნობილი შეფასებები. ამოცანა ბუნებრივად განიხილება სობოლევის შესაბამის სივრცეში. ნახსენები ფუნქციონალის შეფასებად შემოთავაზებულია ე.წ. „ჩასმის“ შეფასება. ნაჩვენებია ძალდებულობისა და ასიმპტოტიურად ნორმალურობის თეორემების სამართლიანობა. დადგენილია კრებადობის რიგები. ზოგადი მეთოდიკა გამოყენებულია რამდენიმე კერძო შემთხვევისათვის. აქ გადაწყვეტილია პრისტლი-ჩაოს რეგრესიის ფუნქციის ფიშერის ინფორმაციული ინტეგრალისა და შენონის ენტროპიის შეფასების პრობლემა.

REFERENCES

1. *M. B. Priestley, M. T. Chao* (1972), Journal of the Royal Statistical Society. Series B (Methodological), **34**, 3: 385-392.
2. *L. Goldstein, K. Messer* (1992), Annals of Statistics, **20**, 3: 1306-1328.
3. *N. Bissantz, H. Holzmann* (2007), J. Statist. Plann. Inference, **137**, 3: 712-719.
4. *E. Gine, D. M. Mason* (2008), Scand. J. Statist. **35**, 4: 739-761.
5. *D. M. Mason, E. Nadaraya, G. Sokhadze* (2010), Nonparametrics and robustness in modern statistical inference and time series analysis: a Festschrift in honor of Professor Jana Jureckova, Inst. Math. Stat. Collect., Inst. Math. Statist., Beachwood, OH, 7: 153-168.
6. *P. K. Babilua, E. A. Nadaraya, M. B. Patsatsia, G. A. Sokhadze* (2008), Proc. I. Vekua Institute of Applied Mathematics, **58**,110: 6-14.
7. *B. Ya. Levit* (1978), Problemy Peredachi Informatsii, **14**, 3: 65-72 (in Russian).
8. *Z. Hlavka* (2011), Acta Universitatis Carolinae - Mathematica et Physica **52**,1: 5-13.
9. *L. Devroye* (1991), NATO ASI Series C. Math. Phys. Sci., **335**: 31-44. Kluwer Acad. Publ., Dortrecht.
10. *J. Kuelbs* (1976), Ec. d'Ete Probab. Saint-Flour V-1975; Lecture Notes in Mathematics, **539**, 224-314. Springer, New York.

Received June, 2014