ASYMPTOTIC BEHAVIOR OF THE SUMMARIZED SQUARE ERROR IN DEPENDENCE DOSE-EFFECT FOR INDIRECT OBSERVATIONS

M.S. Tikhov, D.S. Krishtopenko
Nizhny Novgorod State University
Nizhny Novgorod, RUSSIA
e-mail: tikhovm@mail.ru, krisdima@mail.ru

Abstract

The goal of this paper is to declare results of the asymptotic behavior of the summarized square error of the kernel distribution function estimator $F_n(x)$ defined by $SU_m = \sum_{j=1}^{m}(F_n(x_j) - F(x_j))^2$, where $F(x)$ is the unknown distribution function of a random variable X, $\omega(x)$ is the weight function in dose-response dependence on the sample $\mathcal{U}^{(n)} = \{(W_i, Y_i), 1 \le i \le n\}$, $W_i = I(X_i < U_i)$ is the indicator of even $(X_i < U_i)$ and Y is a random variable, statistically dependant by U. We apply this result to test goodness-of-fit of the distribution function $F(x)$.

1 Introduction

Let $\{(X_i, Y_i, U_i), 1 \le i \le n\}$ be a random sample with a distribution function $F(x)G(y,u)$ and density $f(x)g(y,u)$ on $\mathbf{R}^3$, where $\{X_i\}$ and $\{Y_i, U_i\}$ are independent distributed random variables. We see a sample $\mathcal{U}^{(n)} = \{(W_i, Y_i), 1 \le i \le n\}$, where $W_i = I(X_i < U_i)$ is indicator of event $(X_i < U_i)$. The most nonparametric $\mathcal{U}^{(n)}$-sample estimation of may be written in the form (see [1, 2, 3]).

$$F_n(x) = \frac{S_{2n}(x)}{S_{1n}(x)}, \tag{1}$$

$$S_{1n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(Y_i - x), \quad S_{2n}(x) = \frac{1}{n}\sum_{i=1}^{n} W_i K_h(Y_i - x),$$

and $K(.) \ge 0$ is a *kernel function*, $h > 0$ is a sequence of constants converging to zero as $n \to \infty$ and $K_h(x) = (1/h)K(x/h)$.

In [2] the variable U is treated as inserted for organism dose and X is treated as minimal working dose. Unlike the paper [4] we consider the case, when U is measured as random variable, Y is charactetistic metering error of U.

Define $\| K \|^2 = \int K^2(x)\,dx$, $\nu^2 - \int x^2 K(x)\,dx$, $R(x) = \int F(u)g(|x)\,dx$, $m(y) = \int F(u)g(y,u)\,du$, $q(y) = \int g(y,u)\,du > 0$, $g(u) = \int g(y,u)\,dy > 0$ and

$$g(u|y) = \frac{g(y,u)}{q(y)}, \quad q(y|u) = \frac{g(y,u)}{g(u)},$$

are apparent densities of conformable distributions. And so $R(x) = m(x)q(x)$.

Use the following conditions ($\boldsymbol{K}$).

(**K1**) $K(x) \geq 0$ is a bounded even function on $\mathbf{R}^1$ and $\| K \|^2 < \infty$.

(**K2**) $\int K(x)\,dx = 1$, $\nu^2 = \int x^2 K(x)\,dx < \infty$, $d^4 = \int x^4 K(x)\,dx < \infty$.

(**K3**) $\int x^k K(x)\,dx, k = 1, 3$.

(**K4**) $K(x) = 0$ for $x \notin [-1, 1]$.

(**K5**) $h = h(n)$ is satisfying $h \to 0, nh \to \infty$ as $n \to \infty$.

(**K6**) $f'(x)$ is a continuous function, $\int (f'(x))^2\,dx < \infty$ and is a bounded function on $\mathbf{R}$.

(**K7**) $f(x)/F(x)$, $f'f(x)/F(x)$ are bounded integrand and $\int (f'(x))^4\,dx < \infty$.

Use the following conditions ($\boldsymbol{A}$).

(**A1**) $g(y, u)$ has got 2nd derivatives, which are bounded and continuous function on $\mathbf{R}^2$.

(**A2**) $g(y, u)$ has got 3rd derivatives, which are bounded on $\mathbf{R}^2$.

It is known (see [2]), that on conditions ($\boldsymbol{K}$) $S_{1n}(x) \xrightarrow[n\to\infty]{d} q(x)$ and $S_{2n}(x) \xrightarrow[n\to\infty]{d} m(x)$. Moreover we have the following results.

Theorem 1 [3] . *Under the conditions* (**K**), (**A**) *and* $h = Mn^{-1/5}$, *than*

$$\sqrt{nh}(F_n(x) - R(x)) \xrightarrow[n\to\infty]{d} N(a(x), \sigma^2(x)),$$

$$\text{where} \quad a(x) = M^{5/2} \frac{m''(x)q(x) - q''(x)m(x)}{q^2(x)} \quad \text{and} \quad \sigma^2(x) = \frac{\nu^2 R(x)(1 - R(x))}{q(x)}.$$

The most widely accept measure of the global performance of $F_n(x)$ is their integrated square error $\int (F_n(x) - F(x))^2\,dx$ (see [6]). Here we are considering the summarized square error in indicated points x_j, which are chosen conventionally:

$$SU_m = \sum_{j=1}^{m} (F_n(x_j) - F(x_j))^2. \tag{2}$$

The goal of this paper is to consider the asymptotic behavior of the statistic SU_m.

2 The results

For this purpose we ascertain the asymptotic normality of the multivariate random variable $(F_n(x_1), F_n(x_2), \ldots, F_n(x_m))$.

Theorem 2. *Under the conditions* ($\boldsymbol{K}$) *and* ($\boldsymbol{A}$),

(i) *if* $nh^5 \to \infty$, *than* $\sqrt{nh}(F_n(x_1) - F(x_1), F_n(x_2) - F(x_2), \ldots, F_n(x_m) - F(x_m))$ *is weakly convergent to the normal vector* $(Z_1, Z_2, \ldots, Z_m))$ *of zero means,* $\mathbf{cov}(Z_i, Z_j) = -(1/2)a(x_i)a(x_j), i \neq j,$ *and the variances* $\mathbf{D}(Z_j) = \sigma^2(x_j)$.

(ii) *if* $nh^5 \to 0$, *than* $h^2(F_n(x_1) - F(x_1), F_n(x_2) - F(x_2), \ldots, F_n(x_m) - F(x_m))$ *is weakly convergent to the normal vector* $(Z_1, Z_2, \ldots, Z_m))$ *of zero means, zero covariations and the variances* $\mathbf{D}(Z_j) = \sigma^2(x_j)$.

(iii) *if* $nh^5 \to M \in (0,\infty)$, *than* $n^{-2/5}(F_n(x_1)-F(x_1), F_n(x_2)-F(x_2), \ldots, F_n(x_m)-F(x_m))$ *is weakly convergent to the normal vector* $(Z_1, Z_2, \ldots, Z_m))$ *of zero means, covariations* $\mathbf{cov}(Z_i, Z_j) = -(1/2)a(x_i)a(x_j), i \neq j,$ *and the variances* $\mathbf{D}(Z_j) = \sigma^2(x_j)$.

Define the following matrixes:

$$D = \begin{pmatrix} \sigma^2(x_1) & -(1/2)a(x_1)a(x_2) & \ldots & -(1/2)a(x_1)a(x_2) \\ -(1/2)a(x_2)a(x_1) & \sigma^2(x_1) & \ldots & -(1/2)a(x_2)a(x_m) \\ . & . & \ldots & . \\ -(1/2)a(x_m)a(x_1) & -(1/2)a(x_m)a(x_2) & \ldots & \sigma^2(x_m) \end{pmatrix}$$

and

$$A = \begin{pmatrix} \sigma^2(x_1) & 0 & \ldots & 0 \\ 0 & \sigma^2(x_1) & \ldots & 0 \\ . & . & \ldots & . \\ 0 & 0 & \ldots & \sigma^2(x_m) \end{pmatrix}.$$

The following theorem reveals the asymptotic behavior of the statistic $SU_m = \sum_{j=1}^{m}(F_n(x_j) - F(x_j))^2$.

Theorem 3. *Under the conditions* (***K***) *and* (***A***),

(i) *if* $nh^5 \to \infty$, *than* $\sqrt{nh}\, SU_m$ *is weakly convergent to the random variable with the density*

$$\psi(y) = \frac{my^{m-1}\sqrt{\det D}}{2^{m/2}\Gamma(1+m/2)} \exp(-(1/2)\xi^T D\xi);$$

(ii) *if* $nh^5 \to \infty$, *than* $h^2\, SU_m$ *is weakly convergent to the random variable with the density*

$$\psi(y) = \frac{my^{m-1}\sqrt{\det A}}{2^{m/2}\Gamma(1+m/2)} \exp(-(1/2)\xi^T A\xi);$$

(iii) *if* $nh^5 \to \infty$, *than* $n^{-2/5}\, SU_m$ *is weakly convergent to the random variable with the density*

$$\psi(y) = \frac{my^{m-1}\sqrt{\det D}}{2^{m/2}\Gamma(1+m/2)} \exp(-(1/2)\xi^T D\xi);$$

where components of the vector $\xi = (\xi_1, \xi_2, \ldots, \xi_m)$ *are defined from the condition of standardization and* $\sum_{k=1}^{m} \xi_k^2 = y$.

These results we use for computer testing of statistical homogeneity and goodness of fit hypothesis in this problem.

References

[1] Watson G.S. (1964) Smooth regression analysis. *Sankhya*. Vol. **26**, pp. 359-372.

[2] Tikhov M.S. (2004) Statistical Estimation on the Basis of Interval-Censored Data. *J. Math. Sciences*. Vol. **119**, no 3, pp. 321-335.

[3] Tikhov M.S. (2004) 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., XLIV, pp. 209-215.

[4] Tikhov M.S., Krishtopenko D.S., Yarochuk M.V. (2006) Estimation of the distribution in dependence dose-effect at the fixed plan of experiment. *Statistical methods of estim. and testing statist. hypotheses*. PSU, Perm, pp. 66-77.

[5] Hall P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*. Vol. **14**, pp.1-16.

[6] Hall P. (1984) Integrated square error properties of kernel estimation of regression functions. *Annals of Statistics*. Vol. **12**, no 1, pp. 241-260.