საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 17, №4, 2023
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 17, no. 4, 2023

Mathematics

On the Limit Distribution of Integral Square Deviation between Projection Type Estimators of Distribution Density in $p \geq 2$ Independent Samples

Petre Babilua* **and Elizbar Nadaraya****

* *Department of Mathematics, Faculty of Exact and Natural Sciences, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia*
** *Academy Member, Department of Mathematics, Faculty of Exact and Natural Sciences, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia*

The limiting distribution of the statistic, which describes the mutual deviation of the projection type estimates from each other of distribution density in $p \geq 2$ independent samples is established. The goodness-of-fit test is constructed. Various examples are given. *© 2023 Bull. Georg. Natl. Acad. Sci.*

limit distribution, goodness-of-fit test, distribution density, projection estimator

1. Let $X^{(i)} = \left(X_1^{(i)}, \ldots, X_{n_i}^{(i)}\right)$, $i = 1, \ldots, p$ be independent samples of size $n_1, n_2, \ldots, n_p$, from p $(p \geq 2)$ general populations with distribution densities $f_1(x), \ldots, f_p(x)$. Let, further, $L_2(r)$ be the space of functions with square-integrable measure μ, $d\mu = r(x)dx$ and $\{\varphi_i(x)\}$ be any arbitrary orthonormal basis in this space.

Suppose, that the desired density $f_i(x) \in L_2(r)$, $i = 1, \ldots, p.$ Based on independent samples $X^{(i)}$, $i = 1, \ldots, p,$ construct $p \geq 2$ nonparametric projection estimates for unknown $f_i(x)$.

$$\hat{f}_i(x) = \sum_{j=1}^{\lambda_i(n_i)} \hat{\alpha}_j(i)\varphi_j(x), \quad \hat{\alpha}_j(i) = \frac{1}{n_i}\sum_{k=1}^{n_i} \alpha_j\left(X_k^{(i)}\right),$$
$$\alpha_j(x) = \varphi_j(x)r(x), \quad \lambda_i(n_i) = o(n_i), \quad i = 1, \ldots, p. \tag{1}$$

Projection estimate of distribution density (1) was first introduced and studied by Chencov N. N. [1].

In the present paper, we consider the problem of testing the simple hypothesis, according to which

$$H_0: \quad f_1(x) = f_2(x) = \cdots = f_p(x) \equiv f_0(x)$$

($f_0(x)$ is given density function). For testing this hypothesis we consider criterion of testing hypothesis based on statistic

$$T\left(n_1,n_2,\ldots,n_p\right)=\sum_{i=1}^{P}N_i\int\left[\hat{f}_i(x)-\frac{1}{N}\sum_{j=1}^{P}N_j\hat{f}_j(x)\right]^2 r(x)dx,$$
$$N_i=\frac{n_i}{\lambda_i},\quad i=1,\ldots,p,\quad N=N_1+\cdots+N_p \tag{2}$$

describing the mutual deviation of estimates $\hat{f}_i(x)$, $i=1,\ldots,p$, from each other. In particular case when $p=2$ the statistic T takes more explicit form

$$T(n_1,n_2)=\frac{N_1N_2}{N_1+N_2}\int\left[\hat{f}_1(x)-\hat{f}_2(x)\right]^2 r(x)dx.$$

2. In this section we consider the question concerning the limiting law of the distribution of statistic (2) for the hypothesis H_0, when n_i tends to infinity so that $n_i=k_i\cdot n$, where $n\to\infty$ and k_i are constants. Let $\lambda_1=\lambda_2=\cdots=\lambda_p=\lambda(n)$, where $\lambda(n)\to\infty$ as $n\to\infty$.

Assumptions: $r(x)$, $\varphi_j(x)$, $j=1,2,\ldots$ have bounded variations $V_j<\infty$, $r(x)f_0(x)$ is bounded and $r(x)$ is integrable.

Notations:

$$\Delta_n(f_0)=\frac{1}{\lambda_n}\sum_{j=1}^{\lambda_n}\int\alpha_j^2(x)f_0(x)dx,\quad \lambda_n\equiv\lambda(n),$$

$$\sigma_n^2(f_0)=\frac{2}{\lambda_n}\sum_{i=1}^{\lambda_n}\sum_{j=1}^{\lambda_n}\left(\int\alpha_j(x)\alpha_i(x)f_0(x)dx\right)^2,$$

$$K_n(x,y)=\sum_{j=1}^{\lambda_n}\varphi_j(x)\varphi_j(y)r(y),$$

$$b_n=\sum_{i=1}^{\lambda_n}\gamma_iV_i,\qquad d_n=\sum_{j=1}^{\lambda_n}\gamma_i^2,\qquad \gamma_j=\sup_x\left|\varphi_j(x)\right|,$$

$$S_n(m)=\lambda_n^{-m}\int\cdots\int K_n(t_1,t_2)K_n(t_2,t_3)\cdots K_n(t_m,t_1)\cdot f_0(t_1)\cdots f_0(t_m)r(t_1)\cdots r(t_m)dt_1\cdots dt_m.$$

The following is true.

Theorem. Let $\Delta_n(f_0)=\mu(f_0)+o\left(\lambda_n^{-1/2}\right)$, $\sigma_n^2(f_0)=\sigma^2(f_0)+o\left(\lambda_n^{-1/2}\right)$ as $n\to\infty$ and for all $m\ge 3$,

$$Q_n(m)\equiv\lambda_n^{m-1}S_n(m)=O(1),\quad n\to\infty.$$

If

$$\frac{d_n^{1/2}b_n\ln n}{\sqrt{n}\sqrt{\lambda_n}}\to 0$$

and

$$\frac{b_n^2\ln^2 n}{n\sqrt{\lambda_n}}\to 0\quad\text{as}\quad n\to\infty,$$

then random variable $\lambda_n^{1/2}(T_n-\mu_0)$, $T_n=T\left(n_1,n_2,\ldots,n_p\right)$ has an asymptotically normal distribution with mathematical expectation 0 and variance σ_0^2, where

$$\mu_0 = (p-1)\mu(f_0), \ \sigma_0^2 = 2(p-1)\sigma^2(f_0).$$

Note. The analogous theorem for kernel estimates of distribution density of Rosenblatt-Parzen is proved in work [2].

The theorem allows us to construct the test of asymptotic level α, $0<\alpha<1$ for testing hypothesis H_0, according to which

$$f_1(x) = \cdots = f_p(x) \equiv f_0(x).$$

For this we should calculate T_n and reject H_0, if

$$T_n \ge \mu_0 + \lambda_n^{-1/2}\varepsilon_\alpha \sigma_0^2, \tag{3}$$

where ε_α is the quantile of the level α of a standard normal distribution.

3. Examples of application of theorem.

(a) Let $X_1^{(i)}$, $i=1,\dots,p$ be a bounded random variable

$$c' \le X_1^{(i)} \le c'', \ \varphi_j(x) = \sqrt{\lambda_n (c''-c')^{-1}} \cdot I_j(x) \ [3, 4], \tag{4}$$

where $I_j(x)$ – indicator of interval

$$c' + (j-1)w \le x \le c' + jw, \ w = (c''-c')\cdot \lambda_n^{-1}.$$

In this case $r(x) \equiv 1$, $d_n = O(\lambda_n^2)$ and $b_n = O(\lambda_n^2)$. The conditions $\frac{d_n^{1/2} b_n \ln n}{\sqrt{\lambda_n}\sqrt{n}} \to 0$ and $\frac{b_n^2 \ln^2 n}{n\sqrt{\lambda_n}} \to 0$ are met for $\lambda_n = n^\alpha$, $\alpha < \frac{1}{5}$. Other conditions are also met if $f_0'(x)$ is bounded. It is easy to calculate

$$\Delta_n(f_0) = (c''-c')^{-1},$$

$$\sigma_n^2(f_0) = \frac{2}{c''-c'}\int_{c'}^{c''} f_0^2(x)\,dx + O\left(\frac{1}{\lambda_n}\right) = \sigma^2(f_0) + O\left(\frac{1}{\lambda_n}\right),$$

$$Q_n(m) = \int_{c'}^{c''} \left[f_0(x)\right]^m dx + O\left(\frac{1}{\lambda_n}\right), \ \ m \ge 3.$$

(b) Let $-\pi \le X_1^{(i)} \le \pi$, $i=1,\dots,p$ and $\varphi_j(x)$, $j=1,2,\dots$ – system of trigonometric functions on $[-\pi,\pi]$ ([3, 4]). It is easy to see $d_n = O(\lambda_n)$ and $b_n = O(\lambda_n^2)$. The conditions $\frac{d_n^{1/2} b_n \ln n}{\sqrt{\lambda_n}\sqrt{n}} \to 0$ and $\frac{b_n^2 \ln^2 n}{n\sqrt{\lambda_n}} \to 0$ are met for $\lambda_n = n^\alpha$, $\alpha < \frac{1}{4}$. Further, assuming that $f_0'(x)$ is bounded and use the method of proving of Theorem 3.9 from [5, p. 151] we get

$$\Delta_n(f_0) = \frac{1}{2\pi} + o\left(\frac{1}{\sqrt{\lambda_n}}\right),$$

$$\sigma_n^2(f_0) = \frac{1}{\pi}\int_{-\pi}^{\pi} f_0^2(x)\,dx + o\left(\frac{1}{\sqrt{\lambda_n}}\right),$$

$$\left|Q_n(m)\right| \le c_1 \lambda_n^{-1}(L_n)^m, \ \ m \ge 3$$

and $L_n \sim 4\pi^{-2} \ln \lambda_n$ – Lebesgue constant [5].

(c) Let $-1 \le X_1^{(i)} \le 1,\ i = 1, \ldots, p$ and $\varphi_j(x) = \sqrt{\frac{2j+1}{2}} p_j(x),\ -1 \le x \le 1,\ j = 1, 2, \ldots,$ where $p_j(x)$ are Legendre polynomials [3]. In this case $r(x)dx = dx$. Since

$$\overset{1}{\underset{-1}{\vee}}\left(\varphi_j(x)\right) \le c_2 j^2 \quad \text{and} \quad \sup_x \left|\varphi_j(x)\right| \le \sqrt{2j+1},$$

then

$$d_n = O\left(\lambda_n^2\right) \quad \text{and} \quad b_n^2 = O\left(\lambda_n^{7/2}\right).$$

The conditions $\frac{d_n^{1/2} b_n \ln n}{\sqrt{n}\sqrt{\lambda_n}} \to 0$ and $\frac{b_n^2 \ln^2 n}{n\sqrt{\lambda_n}} \to 0$ are met for $\lambda_n = n^\alpha,\ \alpha < \frac{1}{8}$.

Further, assuming that $f_0(x)$ has bounded derivative, it is easy to get [3], that is

$$\Delta_n(f_0) = \frac{1}{\pi}\int_{-1}^{1} f_0(x)\frac{1}{\sqrt{1-x^2}}dx + O\left(\frac{\ln n}{\lambda_n}\right) = \mu(f_0) + O\left(\frac{\ln n}{\lambda_n}\right),$$

$$\sigma_n^2(f_0) = \frac{2}{\pi}\int_{-1}^{1} f_0^2(x)\frac{1}{\sqrt{1-x^2}}dx + O\left(\frac{1}{\lambda_n}\right) = \sigma^2(f_0) + O\left(\frac{1}{\lambda_n}\right),$$

$$Q_n(m) = \frac{1}{\pi}\int_{-1}^{1} f_0^m(x)\frac{1}{\sqrt{1-x^2}}dx + O(1), \quad m \ge 3.$$

The results obtained in the above examples (a), (b) and (c) can be applied for constructing critical region (3). For example in case (b) critical region (3) will be:

$$T_n \ge (p-1)\frac{1}{2\pi} + 2\varepsilon_\alpha (p-1)\frac{1}{\pi}\int_{-\pi}^{\pi} f_0^2(x)dx, \quad p \ge 2.$$

მათემატიკა

განაწილების სიმკვრივის პროექციული შეფასებების ინტეგრალური კვადრატული გადახრის ზღვარითი განაწილების შესახებ $p \geq 2$ დამოუკიდებელ შერჩევაში

პ. ბაბილუა[*] და ე. ნადარაია[**]

[*] ივანე ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტი, ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი, მათემატიკის დეპარტამენტი, თბილისი, საქართველო
[**] აკადემიის წევრი, ივანე ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტი, ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი, მათემატიკის დეპარტამენტი, თბილისი, საქართველო

მოძებნილია სტატისტიკის ზღვარითი განაწილება, რომელიც აღწერს განაწილების სიმკვრივის პროექციული ტიპის შეფასებების ურთიერთგადახრას $p \geq 2$ დამოუკიდებელ შერჩევაში. განხილულია სხვადასხვა მაგალითი.

REFERENCES

1. Chencov N. N. (1962) Otsenka neizvestnoi plotnosti raspredeleniia po nabliudeniiam, *Dokl. Akad. Nauk SSSR,* **147:** 45-48 (in Russian).
2. Babilua P. K., Nadaraya E. A. (2018) On a homogeneity criterion based on quadratic deviations between kernel estimates for the distribution density in $p \geq 2$ independent samples. *Teor. Veroyatn. i Primen.*, **63**, 4: 654-668 (in Russian).
3. Nadaraya E. A. (1989) Nonparametric estimation of probability densities and regression curves. *Mathematics and its Applications* (Soviet Series), 20. Kluwer Academic Publishers Group, Dordrecht.
4. Nadaraya E. A. (1976) O kvadraticheskoi mere otkloneniia proektsionnoi otsenki plotnosti raspredeleniia, *Teor. Veroyatn. i Primen.*, **21**, 4: 864-871 (in Russian).
5. Zigmund A. (1965) Trigonometricheskie riady, tom **1, 2,** izdatel'stvo "Mir", M. (in Russian).

Received August, 2023