საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 18, №4, 2024
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 18, no. 4, 2024

Mathematics

On the Power of One Goodness-of-Fit Test Based on Square Deviations between Chencov Type Estimators of Distribution Density in $p \geq 2$ Independent Samples

Petre Babilua[*], Elizbar Nadaraya[*,]**

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

A goodness-of-fit test is constructed based on projection type estimates of distribution density. The limiting power of the constructed goodness-of-fit test is stated for Pitman types of "close" alternatives. *© 2024 Bull. Georg. Natl. Acad. Sci.*

Goodness-of-fit test, projection estimator, limit distribution, power

Let $X^{(i)} = \left(X_1^{(i)}, \dots, X_{n_i}^{(i)}\right)$, $i = 1, \dots, p$, be independent samples of size $n_1, n_2, \dots, n_p$, from p $(p \geq 2)$ general populations with distribution densities $f_1(x), \dots, f_p(x)$. 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 complete orthonormal system in this space.

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

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

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 a given density function) against Pitman type "close" alternatives:

$$H_1: \quad f_i(x) = f_0(x) + \alpha(n_0)\psi_i(x), , \quad (2)$$

$$\alpha\left(n_0\right)\to 0,\ \ n_0=\min\left(n_1,\dots,n_p\right)\to\infty,\ \int\psi_i\left(x\right)dx=0,$$
$$\psi_i\left(x\right)\in L_2\left(r\right),\ \ i=1,\dots,p.$$

For testing this hypothesis, we consider criterion of testing hypothesis based on statistic [2]:

$$T\left(n_1,\dots,n_p\right)=\sum_{i=1}^{p}N_i\int\left[\hat f_i\left(x\right)-\frac{1}{N}\sum_{j=1}^{p}N_j\hat f_j\left(x\right)\right]^2 r\left(x\right)dx,$$
$$N_i=\frac{n_i}{\lambda_i},\ \ i=1,\dots,p,\ \ N=N_1+\dots+N_p, \tag{3}$$

describing the mutual deviation of estimates $\hat f_i\left(x\right),\ i=1,\dots,p,$ from each other. In particular case, when $p=2$ the statistic T takes more explicit form

$$T\left(n_1,n_2\right)=\frac{N_1N_2}{N_1+N_2}\int\left[\hat f_1\left(x\right)-\hat f_2\left(x\right)\right]^2 r\left(x\right)dx\,.$$

This particular case was considered in [3].

Let us consider the question about the limiting law of the distribution of statistic (3) for the hypothesis H_1, when n_i tends to infinity so that $n_i=nk_i$, where $n\to\infty$ and k_i are constants. Let $\lambda_1=\lambda_2=\dots=\lambda_p=\lambda\left(n\right)$, where $\lambda\left(n\right)\to\infty$ as $n\to\infty$.

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

Notations:

$$\Delta_n\left(f_0\right)=\frac{1}{\lambda_n}\sum_{j=1}^{\lambda_n}\int\alpha_j^2\left(x\right)f_0\left(x\right)dx,\quad \lambda_n\equiv\lambda\left(n\right),$$
$$\sigma_n^2\left(f_0\right)=\frac{2}{\lambda_n}\sum_{i=1}^{\lambda_n}\sum_{j=1}^{\lambda_n}\left(\int\alpha_i\left(x\right)\alpha_j\left(x\right)f_0\left(x\right)dx\right)^2,$$
$$K_n\left(x,y\right)=\sum_{i=1}^{\lambda_n}\varphi_j\left(x\right)\varphi_j\left(y\right)r\left(y\right),$$
$$d_n=\sum_{i=1}^{\lambda_n}\varphi_iV_i,\qquad \varphi_i=\sup_x\left|\varphi_i\left(x\right)\right|,$$
$$S_n\left(m\right)=\lambda_n^{-m}\int\cdots\int K_n\left(t_1,t_2\right)K_n\left(t_2,t_3\right)\cdots K_n\left(t_m,t_1\right)\cdot f_0\left(t_1\right)\cdots f_0\left(t_m\right)r\left(t_1\right)\cdots r\left(t_m\right)dt_1\cdots dt_m.$$

The following is true.

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

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

Then if there is such a $0<\delta_0<1$ that

$$n^{-1/2}d_n\ln n\to 0,\quad \alpha_nd_n^2\to 0\quad\left(\alpha_n=\alpha\left(n_0\right)\right)$$

for $\lambda_n = n^{\delta}$, $0 < \delta \le \delta_0$ and $\alpha_n = n^{(\delta-2)/4}$, , then for the alternative H_1

$$\lambda_n^{1/2}\left(T_n - \mu_0\right) \xrightarrow{d} N\left(A(\psi), \sigma_0^2\right),$$

where

$$A(\psi) = \sum_{i=1}^{p} k_i \int \left[\psi_i(x) - \frac{1}{\bar{k}} \sum_{j=1}^{p} k_j \psi_j(x) \right]^2 r(x)\, dx,$$

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

$$\bar{k} = k_1 + \cdots + k_p, \quad p \ge 2.$$

d denotes the convergence in distribution and $N\left(a, b^2\right)$ – normally distributed random variable with mathematical expectation a and variance b^2.

Corollary 1. *Let*

$$\Delta_n(f_0) = \mu(f_0) + o\left(\lambda_n^{-1/2}\right),$$

$$\sigma_n^2(f_0) = \sigma^2(f_0) + o\left(\lambda_n^{-1/2}\right) \quad as \quad n \to \infty,$$

and for all $m \ge 3$

$$Q_n(m) = \lambda_n^{m-1} S_n(m) = O(1).$$

If

$$\frac{d_n \ln n}{\sqrt{n}} \to 0,$$

then random variable $\lambda_n^{1/2}\left(T_n - \mu_0\right)$ for the hypothesis H_0 has normal distribution $N\left(0, \sigma_0^2\right)$ [2].

Based on corollary, we can construct criterion for testing hypothesis H_0. Critical domain for testing this hypothesis can be established by the inequality

$$T_n \ge d_n(\alpha), \tag{4}$$

where $d_n(\alpha) = \mu_0 + \lambda_n^{-1/2}\sigma_0 \varepsilon_\alpha$, ε_α, ε_α is the quantile of the level $1-\alpha$ $(0 < \alpha < 1)$ of a standard normal distribution $\Phi(x)$.

Corollary 2. Under conditions of theorem local behavior of the power $P_{H_1}\left(T_n \ge d_n(\alpha)\right)$ is as follows

$$P_{H_1}\left(T_n \ge d_n(\alpha)\right) \longrightarrow 1 - \Phi\left(\varepsilon_\alpha - \frac{A(\psi)}{\sigma_0}\right),$$

where

$$A(\psi) = \sum_{i=1}^{p} k_i \int \left[\psi_i(x) - \frac{1}{\bar{k}} \sum_{j=1}^{p} k_j \psi_j(x) \right]^2 r(x)\, dx,$$

$$\bar{k} = k_1 + k_2 + \cdots + k_p, \quad p \ge 2.$$

It should be noted that criterion (4) for testing hypothesis H_0 against alternatives type (2) is asymptotically strictly unbiased, since $A(\psi) > 0$, and is equal to 0 if and only if

$$\psi_1(x) = \psi_2(x) = \cdots = \psi_p(x).$$

Example. 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 $[-\pi,\pi]$. It is easy to see $d_n = O\left(\lambda_n^2\right)$. The conditions $\dfrac{d_n \ln n}{\sqrt{n}} \to 0,\ \alpha_n d_n^2 \to 0$ for $\lambda_n = n^{\delta},\ \alpha_n = n^{-1/2+\delta/4}$ are met if $0 < \delta < \delta_0 = \dfrac{2}{17}$.

Further, assuming that $f_0'(x)$ is bounded and use the method of proving of Theorem 3.9 from [4: 151], we get

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

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

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

and $L_n \sim 4\pi^{-2} \ln \lambda_n$ – Lebesgue constant [4]. In this case critical domain (4) will be

$$T_n \ge (p-1)\cdot \frac{1}{2\pi} + \left[2(p-1)\right]^{1/2} \varepsilon_\alpha \cdot \lambda_n^{-1/2} \left(\frac{1}{\pi}\int_{-\pi}^{\pi} f_0^2(x)\,dx\right)^{1/2}.$$

მათემატიკა

განაწილების სიმკვრივის ჩენცოვის ტიპის შეფასებების კვადრატულ გადახრებზე დაფუძნებული ერთი თანხმობის კრიტერიუმის სიმძლავრის შესახებ $p \geq 2$ დამოუკიდებელ შერჩევაში

პ. ბაბილუა[*], ე. ნადარაია[*,]**

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

აგებულია განაწილების სიმკვრივის პროექციულ შეფასებებზე დაფუძნებული თანხმობის კრიტერიუმი. მოძებნილია აგებული თანხმობის კრიტერიუმის ზღვარითი სიმძლავრე პიტმანის ტიპის დაახლოებადი ალტერნატივებისთვის.

REFERENCES

1. Chencov N. N. (1962) A bound for an unknown distribution density in terms of the observations. Dokl. Akad. Nauk SSSR **147:** 45-48.
2. Babilua P. K., Nadaraya E. A. (2023) On the limit distribution of integral square deviation between projection type estimators of distribution density in $p \geq 2$ independent samples. *Bull. Georg. Natl. Acad. Sci.,* (N.S.) 17, 4: 15-19.
3. Mirzakhmedov M. A. (1982) Nonparametric estimation of a probability density by orthonormal functions. *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk,* 4: 21-25 (in Russian).
4. Zigmund A. (1965) Trigonometricheskie serii (Trigonometric Series). Vols. I, II. Izdat. “Mir“, M. (in Russian).

Received October, 2024