M. GUTMAN (Wrocław)

O ESTYMOWANIU CAŁKI KWADRATU GĘSTOŚCI

1. H. Steinhaus zwrócił uwagę na to, że całkę kwadratu gęstości prawdopodobieństwa można używać jako miarę skupienia rozkładu i że w licznych przypadkach, dzięki swoim własnościom, szczególnie się do tego celu nadaje [3]. W związku z tym powstaje zagadnienie estymacji tego parametru. W niniejszej nocie ograniczymy się jedynie do przypadku, gdy $f(x)$ jest gęstością prawdopodobieństwa rozkładu normalnego. Podamy nieobciążony estymator całki

$$\int_{-\infty}^{\infty} f^2(x)\,dx, \tag{1}$$

to jest estymator, którego wartość oczekiwana jest równa wartości parametru estymowanego. Zagadnienie estymacji całki (1), gdy $f(x)$ jest gęstością prawdopodobieństwa rozkładu normalnego, jest równoważne z zagadnieniem estymacji precyzji h tego rozkładu, to jest odwrotności odchylenia średniego. Całka (1) jest bowiem w tym przypadku wprost proporcjonalna do h i dlatego zagadnienie to w gruncie rzeczy sprowadza się do podania estymatora nieobciążonego precyzji. W praktyce często bierzemy za estymator precyzji odwrotność nieobciążonego estymatora odchylenia średniego. Okazuje się jednak, że nie jest to nieobciążony estymator precyzji. Wykażemy ponadto, że estymator proponowany przez nas jest nie tylko nieobciążony, lecz również asymptotycznie najefektywniejszy ([2], § 10.5).

2. TWIERDZENIE 1. *Jeśli*

1° *zmienna losowa X ma rozkład normalny z gęstością prawdopodobieństwa*

$$f(x) = e^{-x^2/2\sigma^2}/\sqrt{2\pi}\sigma, \tag{2}$$

2° *średnie odchylenie s jest równe*

$$s = \sqrt{n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

gdzie $x_1, x_2, \ldots, x_n$ *jest n-elementową próbką wylosowaną z normalnej populacji generalnej,*

3° *liczba elementów w próbce jest większa od trzech* ($n > 3$), *to wyrażenie* L_n/s, *gdzie*

$$L_n = \frac{1}{2\sqrt{\pi}}\sqrt{\frac{2}{n}}\cdot\frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\Gamma\left(\frac{1}{2}(n-2)\right)}, \tag{3}$$

jest estymatorem nieobciążonym całki (1).

Dowód. Jak zaznaczyliśmy na wstępie, całka (1), gdy $f(x)$ spełnia wzór (2), jest wprost proporcjonalna do współczynnika precyzji rozkładu, tj. do $1/\sigma$. Rzeczywiście, biorąc pod uwagę (2) mamy

$$\int_{-\infty}^{\infty} f^2(x)\,dx = \int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi}\,\sigma}e^{-x^2/2\sigma^2}\right)^2 dx = \frac{1}{\sqrt{2\pi}\,\sigma}\cdot\frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\cdot\frac{1}{2}\sigma^2}}e^{-x^2/2(\sigma^2/2)}dx.$$

Ostatnia całka jest równa 1, mamy zatem

$$\int_{-\infty}^{\infty} f^2(x)\,dx = \frac{1}{2\sqrt{\pi}}\cdot\frac{1}{\sigma}. \tag{4}$$

Aby udowodnić tezę, wystarczy więc wykazać, że

$$\mathrm{E}\left(\frac{L_n}{s}\right) = \frac{1}{2\sqrt{\pi}}\cdot\frac{1}{\sigma}. \tag{5}$$

Znając gęstość $\varphi(s)$ prawdopodobieństwa zmiennej losowej s,

$$\varphi(s) = \frac{n^{(n-1)/2}e^{-ns^2/2\sigma^2}s^{n-2}}{2^{(n-3)/2}\Gamma\left(\frac{1}{2}(n-1)\right)\sigma^{n-1}} \tag{6}$$

([2], wzór 7.6.2), oznaczając

$$A = n^{(n-1)/2}/2^{(n-3)/2}\Gamma\left(\tfrac{1}{2}(n-1)\right)\sigma^{n-1} \tag{7}$$

i uwzględniając równość

$$\mathrm{E}(L_n/s) = L_n\mathrm{E}(1/s) = L_n\int_0^{\infty}(1/s)\varphi(s)\,ds,$$

znajdujemy

$$L_n\mathrm{E}(1/s) = AL_n\int_0^{\infty}e^{-ns^2/2\sigma^2}s^{n-3}\,ds. \tag{8}$$

Obliczamy całkę

$$I = \int_0^{\infty}e^{-ns^2/2\sigma^2}s^{n-3}\,ds.$$

W tym celu podstawiamy

$$\frac{ns^2}{2\sigma^2} = u, \quad s = \left(\frac{2\sigma^2 u}{n}\right)^{1/2}, \quad ds = \frac{\sigma^2}{ns} du,$$

co daje

$$I = e^{-u}\left(\frac{2\sigma^2 u}{n}\right)^{(n-3)/2} \frac{\sigma^2}{n}\left(\frac{2\sigma^2 u}{n}\right)^{-1/2} du = 2^{(n-4)/2} n^{(-n+2)/2} \sigma^{n-2} \int_0^\infty e^{-u} u^{(n-4)/2} du.$$

Ponieważ

$$\int_0^\infty e^{-u} u^{(n-2)/2-1} = \Gamma\left(\tfrac{1}{2}(n-2)\right),$$

więc

$$I = 2^{(n-4)/2} n^{(-n+2)/2} \sigma^{n-2} \Gamma\left(\tfrac{1}{2}(n-2)\right). \tag{9}$$

Z wzorów (3), (7), (8), (9) otrzymujemy

$$L_n \mathrm{E}(1/s) = AL_n I = (1/2\sqrt{\pi})(1/\sigma), \qquad \text{c. n. d.}$$

Udowodnimy teraz

TWIERDZENIE 2. *Wyrażenie L_n/s jest asymptotycznie najefektywniejszym estymatorem całki* (1).

D o w ó d. Zwróćmy uwagę, że efektywność (oznaczmy ją przez e) rozważanego estymatora jest równa efektywności estymatora nieobciążonego U parametru h, przy czym na mocy (5)

$$U = 2\sqrt{\pi} L_n/s. \tag{10}$$

Wystarczy więc wykazać, że $\mathrm{e}(U) \underset{n\to\infty}{\to} 1$.

Oznaczmy przez $\mathrm{D}^2(\psi)$ wariancję najefektywniejszego estymatora parametru h. Jak wiadomo, efektywność estymatora U jest równa

$$\mathrm{e}(U) = \mathrm{D}^2(\psi)/\mathrm{D}^2(U).$$

Wykażemy najpierw, że

$$\mathrm{D}^2(\psi) = h^2/2n. \tag{11}$$

Rzeczywiście, na podstawie nierówności Rao i Craméra ([1], wzór 10.5.1) mamy

$$\mathrm{D}^2(\psi) = \left(n \int_{-\infty}^{\infty} \left[\frac{\partial \log f(x)}{\partial h}\right]^2 f(x)\,dx\right)^{-1}, \tag{12}$$

gdzie $f(x) = e^{-h^2x^2/2}h/\sqrt{2\pi}$. Logarytmując ostatnią równość i następnie różniczkując względem h, otrzymujemy

$$\partial \log f(x)/\partial h = 1/h - hx^2.$$

Korzystając z ostatniej równości i równości (12), uzyskujemy

$$D^2(\psi) = \left(n \int_{-\infty}^{\infty} (1/h^2 - 2x^2 + h^2x^4) f(x)\,dx\right)^{-1},$$

a stąd

$$D^2(\psi) = \left(n(1/h^2 - 2/h^2 + 3h^2/h^4)\right)^{-1} = h^2/2n,$$

czyli równość (11).

Wykażemy z kolei, że wariancja estymatora U jest równa

$$D^2(U) = h^2\left(\frac{n-3}{2}\cdot\frac{\Gamma^2\left(\frac{1}{2}(n-3)\right)}{\Gamma^2\left(\frac{1}{2}(n-2)\right)} - 1\right). \tag{13}$$

Istotnie, z definicji mamy

$$D^2(U) = \int_0^{\infty} U^2\varphi(s)\,ds - \left(E(U)\right)^2, \tag{14}$$

a pierwsza całka na mocy (10) jest równa

$$\int_0^{\infty} U^2\varphi(s)\,ds = 4\pi L_n^2 \int_0^{\infty} (1/s^2)\varphi(s)\,ds.$$

Przez całkowanie można wykazać, że

$$\int_0^{\infty} (1/s^2)\varphi(s)\,ds = nh^2/(n-3). \tag{15}$$

Rzeczywiście, korzystając z wzorów (6) i (7) mamy

$$\int_0^{\infty} (1/s^2)\varphi(s)\,ds = A\int_0^{\infty} e^{-ns^2/2\sigma^2}s^{-2}s^{n-2}\,ds = A\int_0^{\infty} e^{-ns^2/2\sigma^2}s^{n-4}\,ds.$$

Wyznaczamy całkę

$$I = \int_0^{\infty} e^{-ns^2/2\sigma^2}s^{n-4}\,ds.$$

W tym celu podstawiamy

$$\frac{ns^2}{2\sigma^2} = u, \qquad s = \left(\frac{2\sigma_2 u}{n}\right)^{1/2}, \qquad ds = \frac{\sigma^2}{ns}\,du.$$

Otrzymujemy stąd

$$I = \int_0^\infty e^{-u}\left(\frac{2\sigma^2 u}{n}\right)^{(n-4)/2}\frac{\sigma^2}{n}\left(\frac{2\sigma^2 u}{n}\right)^{-1/2} du = 2^{(n-5)/2} n^{(-n+3)/2}\sigma^{n-3}\int_0^\infty e^{-u} u^{(n-3)/2-1} du.$$

Ostatnia całka jest równa $\Gamma\left(\frac{1}{2}(n-3)\right)$, mamy zatem

$$\int_0^\infty (1/s^2)\varphi(s)ds = AI = \big(n/(n-3)\big)(1/\sigma^2), \qquad \text{c. n. d.}$$

Z kolei, na mocy (10) i (5), mamy

$$(16) \qquad \big(\mathrm{E}(U)\big)^2 = h^2.$$

Z równości (14), (15), (16) otrzymujemy

$$(17) \qquad \mathrm{D}^2(U) = 4\pi L_n^2 \frac{nh}{n-3} - h^2 = h^2\left[\frac{2}{n-3}\cdot\frac{\Gamma^2\left(\frac{1}{2}(n-1)\right)}{\Gamma^2\left(\frac{1}{2}(n-2)\right)} - 1\right].$$

Stosując teraz znaną równość

$$\tfrac{1}{2}(n-3)\Gamma\left(\tfrac{1}{2}(n-3)\right) = \Gamma\left(\tfrac{1}{2}(n-1)\right)$$

o równości (17), otrzymujemy wariancję (13).

Obliczmy teraz efektywność estymatora U. Korzystając z wzorów (11) i (12), uzyskujemy równość

$$\mathrm{e}(U) = \frac{\mathrm{D}^2(\psi)}{\mathrm{D}^2(U)} = h^2\Big/\left(2h^2 n\tfrac{1}{2}(n-3)\frac{\Gamma^2\left(\frac{1}{2}(n-3)\right)}{\Gamma^2\left(\frac{1}{2}(n-2)\right)} - 1\right).$$

Wreszcie, korzystając z równości

$$\left(\frac{n}{2}\cdot\frac{\Gamma^2\left(\frac{1}{2}n\right)}{\Gamma^2\left(\frac{1}{2}(n-1)\right)} - 1\right)\sigma^2 = \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right)$$

(zob. [1], str. 526), otrzymujemy związek

$$\mathrm{e}(U) = \frac{1}{2n/2(n-3)+O(1/n)} \underset{n\to\infty}{\to} 1,$$

co kończy dowód twierdzenia 2.

3. Poniżej podajemy tablice przedstawiające: 1° obliczone współczynniki L_n dla różnych n oraz 2° efektywności estymatora U dla różnych n.

TABLICA 1

Wartości współczynnika L_n estymatora $U = L_n\cdot(1/s)$

n	L_n
3	0,14
5	0,20
7	0,23
25	0,27
100	0,28
∞	0,283

TABLICA 2

Efektywność estymatora $U = L_n\cdot(1/s)$

n	e(U)
4	0,20
10	0,70
15	0,79
51	0,93
101	0,96
∞	1,00

4. Podamy tu jeszcze kilka wniosków wynikających z twierdzeń przez nas podanych.

(a) Estymator

$$U = \sqrt{\frac{2}{n}}\cdot\frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\Gamma\left(\frac{1}{2}(n-2)\right)}\cdot\frac{1}{s}$$

jest nieobciążonym estymatorem współczynnika precyzji h rozkładu normalnego. Wynika to natychmiast z wzorów (2) i (9).

(b) Estymator $U_2 = (n-3)/ns^2$ jest estymatorem nieobciążonym kwadratu współczynnika precyzji h^2 rozkładu normalnego. Wniosek ten wynika z wzoru (15).

(c) Jeśli: 1° zmienne losowe X i Y są niezależne i mają rozkład normalny, 2° σ_x i σ_y są odpowiednio odchyleniami średnimi zmiennych losowych X i Y, 3° s_1 i s_2 są odpowiednio ich średnimi odchyleniami obliczonymi z próbek o liczebnościach n_1 i n_2, to estymator

$$U_3 = \frac{2}{4\pi}\cdot\frac{1}{n_1 n_2}\cdot\frac{\Gamma\left(\frac{1}{2}(n_1-1)\right)\Gamma\left(\frac{1}{2}(n_2-1)\right)}{\Gamma\left(\frac{1}{2}(n_1-2)\right)\Gamma\left(\frac{1}{2}(n_2-2)\right)}\cdot\frac{1}{s_1 s_2}$$

jest estymatorem nieobciążonym całki

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^2(xy)\,dx\,dy.$$

Rzeczywiście, na podstawie założenia niezależności X i Y oraz wzoru (4) mamy

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^2(xy)\,dx\,dy = \frac{h_1}{2\sqrt{\pi}}\cdot\frac{h_2}{2\sqrt{\pi}}.$$

Korzystając dalej z wniosku (a) otrzymamy, że wartość oczekiwana U_3 jest równa $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^2(xy)\,dx\,dy$, c. n. d.

(d) Z wniosku (c) wynika natychmiast, że w przypadku gdy $n = n_1 = n_2$, estymator nieobciążony całki kwadratu gęstości jest równy

$$\operatorname{est}_{\mathrm{nb}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^2(xy)\,dx\,dy = \frac{1}{2\pi n}\cdot\frac{\Gamma^2\left(\frac{1}{2}(n-1)\right)}{\Gamma^2\left(\frac{1}{2}(n-2)\right)}\cdot\frac{1}{s_1 s_2}.$$

Prace cytowane

[1] H. Cramér, *Mathematical methodes of statistics*, Princeton 1946.

[2] M. Fisz, *Rachunek prawdopodobieństwa i statystyka matematyczna*, Warszawa 1954.

[3] H. Steinhaus, *Über einige prinzipielle Fragen der mathematischen Statistik*, Bericht über die Tagung Wahrscheinlichkeitsrechnung u. math. Statistik, Berlin 1954, str. 59-60.

Praca wpłynęła 25. 10. 1956

М. ГУТМАН (Вроцлав)

ОБ ОЦЕНКЕ ИНТЕГРАЛА С КВАДРАТА ПЛОТНОСТИ

РЕЗЮМЕ

Х. Штайнхауз обратил внимание на то, что интеграл с квадрата плотности можно употреблять как меру совокупности распределения и что во многих случаях, благодаря своим свойствам, особенно к этому пригоден.

Оказывается, что если

1° x имеет нормальное распределение с плотностью вероятности определенной по формуле (1),

2° среднее отклонение s равно

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2},$$

причём $x_1, x_2, \ldots, x_n$ является n-элементной выборкой полученной с нормальной генеральной совокупности, тогда оценки

$$\frac{1}{2\sqrt{\pi}}\sqrt{\frac{2}{n}}\cdot\frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\Gamma\left(\frac{1}{2}(n-2)\right)}\cdot\frac{1}{s}$$

является несмещенной и асимптотически наиболее эффективной оценкой интеграла

$$\int_{-\infty}^{\infty} f^2(x)\,dx.$$

С доказанных в статье теорем следует, что оценка $(n-3)/ns^2$ является несмещенной оценкой квадрата точности h^2 нормального распределения.

M. GUTMAN (Wrocław)

ON ESTIMATING AN INTEGRAL FROM THE SQUARE OF THE DENSITY

SUMMARY

H. Steinhaus has pointed out that an integral from the square of probability density may be used as measure of the concentration of a distribution and that in numerous cases, owing to its properties, it is particularly suitable for this purpose.

In this paper the author discusses the problem of estimating that parameter, limiting himself to the case where $f(x)$ is the probability density of the normal distribution.

We find that if

1° the variable x has the normal distribution with the probability density defined by formula (1),

2° the mean deviation s is

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

where $x_1, x_2, \ldots, x_n$ is an n-element sample drawn from a normal general population, then the estimator

$$\frac{1}{2\sqrt{\pi}}\sqrt{\frac{2}{n}}\cdot\frac{\Gamma(\frac{1}{2}(n-1))}{\Gamma(\frac{1}{2}(n-2))}\cdot\frac{1}{s}$$

is an unbiased and asymptotically most effective estimator of the integral $\int_{-\infty}^{\infty} f^2(x)\,dx$.

The theorems contained in the paper imply also that the estimator $(n-3)/ns^2$ is an unbiased estimator of the precision square h^2 of the normal distribution.