Liet. matem. rink., T. 43, spec. nr., 560–564

Statistikos $\rho_N = |Z_N W_N|$ optimalus konvergavimo greičio įvertis

Robertas VILKAS (MII, KTU), Algimantas AKSOMAITIS (KTU)
el. paštas: robertas.vilkas@ktu.lt

1. Įvadas

Tarkime, kad $(X_1, \ldots, X_N)$ yra atsitiktinė imtis, kurioje X_j yra tarpusavyje nepriklausomi ir pasiskirstę tolygiai intervale $(-1, 1)$. Atsitiktinis imties tūris N nepriklauso nuo X_j $(1 \leqslant j \leqslant n)$ ir yra pasiskirstęs pagal geometrinį dėsnį:

$$P(N = k) = \frac{1}{n}\Big(1 - \frac{1}{n}\Big)^{k-1}, \quad k \geqslant 1 \ \text{(iš čia aišku, kad turi būti } n \geqslant 2\text{)}.$$

Pažymėkime

$$Z_N = \max(X_1, \ldots, X_N), \quad W_N = \min(X_1, \ldots, X_N).$$

Šiame straipsnyje pateiksime statistikos $\rho_N = |Z_N W_N|$ pasiskirstymo funkcijos (normuotos) konvergavimo greičio įvertį.

H.M. Barakat ir E.M. Nigm straipsnyje [1] yra suformuluotos ir įrodytos bendrosios ribinės teoremos tokioms netiesinėms statistikoms:

$$\rho_N = |Z_N W_N|, \ \gamma_N = \sqrt{\rho_N} \ \text{ ir } q_N = \frac{W_N}{Z_N}.$$

Ten yra gauti visi galimi ribiniai skirstiniai (atliekant tiesinį normavimą) ir netgi užrašytos normavimo konstantos.

Straipsnyje [1] yra pateikta keletas pavyzdžių (tuose pavyzdžiuose N pasiskirstęs pagal geometrinį dėsnį). Mūsų tiriamas atvejis (kai $X_j \sim T(-1, 1)$) sutampa su vienu iš tų pavyzdžių. Tas atvejis buvo įdomus todėl, kad tame pavyzdyje buvo rasta klaida (žr. žemiau), kurią čia ištaisėme.

2. Statistikos $\rho_N = |Z_N W_N|$ skirstinio funkcija

Atsitiktinių dydžių X_j pasiskirstymo funkcija $F(x) = \frac{x+1}{2}$, $-1 < x < 1$. Tuomet vektoriaus (Z_n, W_n) (n-determinuotas) pasiskirstymo tankis

$$f_{Z,W}(x, y) = \begin{cases} 0, & y > x, \\ n(n-1)\,(F(x) - F(y))^{n-2} f(x) f(y), & y \leqslant x, \end{cases}$$

$$= \begin{cases} 0, & y > x, \\ \frac{n(n-1)}{4}\left(\frac{x-y}{2}\right)^{n-2}, & y \leqslant x. \end{cases}$$

Determinuotu atveju ρ_n pasiskirstymo funkcija

$$F_{\rho_n}(z) = P(\rho_n < z) = P(|Z_n W_n| < z) = \iint\limits_{|x\cdot y|<z} f_{Z,W}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

$$= \int_{-\sqrt{z}}^{-z} f_1\,\mathrm{d}x + \int_{-z}^{z} f_2\,\mathrm{d}x + \int_{z}^{\sqrt{z}} f_3\,\mathrm{d}x + \int_{\sqrt{z}}^{1} f_4\,\mathrm{d}x, \quad 0 < z < 1;$$

čia

$$f_1 = \int_{z/x}^{x} \frac{n(n-1)}{4}\left(\frac{x-y}{2}\right)^{n-2}\mathrm{d}y = \frac{n}{2}\left(\frac{x^2-z}{2x}\right)^{n-1},$$

$$f_2 = \int_{-1}^{x} \frac{n(n-1)}{4}\left(\frac{x-y}{2}\right)^{n-2}\mathrm{d}y = \frac{n}{2}\left(\frac{x+1}{2}\right)^{n-1},$$

$$f_3 = \int_{-z/x}^{x} \frac{n(n-1)}{4}\left(\frac{x-y}{2}\right)^{n-2}\mathrm{d}y = \frac{n}{2}\left(\frac{x^2+z}{2x}\right)^{n-1},$$

$$f_4 = \int_{-z/x}^{z/x} \frac{n(n-1)}{4}\left(\frac{x-y}{2}\right)^{n-2}\mathrm{d}y = \frac{n}{2}\left(\frac{x^2+z}{2x}\right)^{n-1} - \frac{n}{2}\left(\frac{x^2-z}{2x}\right)^{n-1}$$
$$= f_3 - f_1.$$

Kai N yra pasiskirstęs pagal geometrinį dėsnį, statistikos ρ_N pasiskirstymo funkcija

$$F_{\rho_N}(z) = \sum_{k=1}^{\infty} P(\rho_k < z)\cdot P(N=k) = \sum_{k=1}^{\infty} P(\rho_k < z)\cdot\frac{1}{n}\left(1-\frac{1}{n}\right)^{k-1}.$$

Pasirodo, kad $F_{\rho_N}(z)$ yra elementarioji funkcija, tačiau gaunama gana gremėzdiška jos išraiška (dėl straipsnio apimtiems ribotumo tarpinių veiksmų nepateikiame). Todėl įveskime tokius pažymėjimus:

$$I_{1,n}(x,z) = -n(n-1)\cdot z\left((n-1)^2 z + n^2\right)^{-\frac{3}{2}}\cdot \operatorname{arth}\left(\frac{(n-1)x-n}{\sqrt{(n-1)^2 z + n^2}}\right)$$
$$- \frac{n(((n-1)^2 z + 2n^2)x + n^2 z - nz)}{(n-1)((n-1)^2 z + n^2)((n-1)(x^2-z) - 2xn)},$$

$$\text{čia } \operatorname{arth}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right);$$

$$I_{2,n}(x) = \frac{2n}{(n-1)(n+1-(n-1)x)};$$

$$I_{3,n}(x,z) = n(n-1)\cdot z((n-1)^2 z - n^2)^{-\frac{3}{2}}\cdot \operatorname{arctg}\left(\frac{(n-1)x-n}{\sqrt{(n-1)^2 z - n^2}}\right)$$

$$-\frac{n(((n-1)^2z-2n^2)x+n^2z-nz)}{(n-1)((n-1)^2z-n^2)((n-1)(x^2+z)-2xn)};$$
$$I_{4,n}(x,z)=I_{3,n}(x,z)-I_{1,n}(x,z).$$

Tuomet pasiskirstymo funkciją galime užrašyti taip:

$$F_{\rho_N}(z) = I_{1,n}(-z,z)-I_{1,n}(-\sqrt{z},z)+I_{2,n}(z)-I_{2,n}(-z)+I_{3,n}(\sqrt{z},z)$$
$$-I_{3,n}(z,z)+I_{4,n}(1,z)-I_{4,n}(\sqrt{z},z), \quad 0<z<1.$$

3. Konvergavimo greitis

Minėtame straipsnyje (H.M. Barakat, E.M. Nigm) teigiama, jog normuota pasiskirstymo funkcija $F_{\rho_N}(\frac{2x}{n}+1)$ artės prie neišgimusios ribinės pasiskirstymo funkcijos $Q^{**}(x;1)$. Čia bendros funkcijos $Q^{**}(x;\alpha)$ išraiškos nepateiksime ($Q^{**}(x;\alpha)$ yra tam tikras dvilypis integralas, priklausantis nuo X_j pasiskirstymo). Taip pat teigiama, jog

$$Q^{**}(x;1)=\begin{cases}\frac{1}{1-x}, & x<0,\\ 1, & x\geqslant 0.\end{cases}$$

Pasirodo, kad pastarasis teiginys yra neteisingas. Iš tikrųjų, paskaičiavus gautume, kad

$$Q^{**}(x;1)=\begin{cases}1-\left(\frac{x}{1-x}\right)^2, & x<0,\\ 1, & x\geqslant 0.\end{cases}$$

Skaičiuojame ribą (tiesiogiai, remiantis mūsų gauta $F_{\rho_N}(z)$ išraiška):

$$L_0(x):=\lim_{n\to\infty}F_{\rho_N}\Big(\frac{2x}{n}+1\Big)=1-\Big(\frac{x}{1-x}\Big)^2, \quad x<0$$
$$\bigg(\text{čia } x<0, \text{ nes } \Big(0<\frac{2x}{n}+1<1\Big)\Rightarrow\Big(-\frac{n}{2}<x<0\Big)\bigg).$$

Gavome tokią pačią ribinę funkciją, t.y., $L_0(x)=Q^{**}(x;1)$. Ribos skaičiavimo dėl straipsnio apimties ribotumo ir funkcijos gremėzdiškumo nepateikiame, bet šią ribą nesunkiai, nes turime elementarią funkciją, galima surasti daugumos simbolinių matematinių paketų pagalba, pavyzdžiui, MAPLE.

Įvertinsime konvergavimo greitį $\Delta_n(x)=F_{\rho_N}(\frac{2x}{n}+1)-L_0(x)$. Aišku, mus domina tik intervalas $-\frac{n}{2}<x<0$, nes už šio intervalo ribų konvergavimo greitis nepriklauso nuo n. Skaičiuojame ribą:

$$L_1(x):=\lim_{n\to\infty}n\cdot\bigg(F_{\rho_N}\Big(\frac{2x}{n}+1\Big)-L_0(x)\bigg)=\frac{x^2(5x+3)}{3(1-x)^3}.$$

Pažymėkime aibes $\mathbf{U}=\{x\colon x<0\}$, $\mathbf{A}=\{(x,n)\colon -\frac{n}{2}<x<0,\ n\in N\backslash\{1\}\}$. Funkcija $L_1(x)$ nėra tapatingai lygi nuliui ir yra aprėžta $\forall x\in\mathbf{U}$, todėl konvergavimo greičio eilė yra $\frac{1}{n}$.

Parodysime, kad $\sup_{(x,n)\in\mathbf{A}} |n \cdot \Delta_n(x)| < \infty$. Pažymėkime

$$a_1 = \inf_{(x,n)\in\mathbf{A}} n \cdot \Delta_n(x), \quad b_1 = \sup_{(x,n)\in\mathbf{A}} n \cdot \Delta_n(x).$$

Kadangi $-\frac{n}{2} < x < 0$ ir $n \geqslant 2$, tai $n \geqslant \max\{2, -2x\}$. Sandaugos $n \cdot \Delta_n(x)$ išvestinė atžvilgiu n yra teigiama, kai $n > \max\{2, -2x\}$ (dėl straipsnio apimties ribotumo įrodymo nepateikiame). Todėl $n \cdot \Delta_n(x)$ yra monotoniškai didėjanti funkcija. Vadinasi, funkcijos $n \cdot \Delta_n(x)$ supremumas bus gaunamas tik tuomet, kai $n = \infty$. Kadangi $\lim_{n\to\infty} n \cdot \Delta_n(x) = L_1(x)$ ir funkcija $L_1(x) = \frac{x^2(5x+3)}{3(1-x)^3}$ aibėje $\mathbf{U}$ turi maksimumą $\frac{1}{48}$ (taške $x = -\frac{1}{3}$), tai $b_1 = \frac{1}{48}$. Dabar raskime infimumą. Jis bus gaunamas, kai $n = \max\{2, -2x\}$. Taigi, yra du atvejai: 1) $n = 2$ ir $-1 < x < 0$; 2) $n = -2x$ ir $x < -1$.

Pirmas atvejis. Funkcijos $\Delta_2(x)$ išvestinė yra teigiama. Todėl mažiausią reikšmę gausime, kai $x = -1$. Skaičiuojame ribą $\lim_{x\to -1+0} \Delta_2(x) = -\frac{3}{8}$. Vadinasi, sandaugos $n \cdot \Delta_n(x)$ mažiausia reikšmė yra $-\frac{3}{4}$ (kai $-1 < x < 0$ ir $n = 2$).

Antras atvejis. Skaičiuojame ribą $\lim_{n\to -2\cdot x+0} n \cdot \Delta_n(x) = -\frac{2(2x^2-x)}{(1-x)^2}$. Pastaroji funkcija yra monotoniškai didėjanti, taigi, infimumas bus gaunamas, kai $x = -\infty$. Skaičiuojame ribą $\lim_{x\to -\infty} -\frac{2(2x^2-x)}{(1-x)^2} = -4$. Taigi, $a_1 = \min\{-\frac{3}{8}, -4\} = -4$. Tokiu būdu įrodėme 1 teoremą.

1 teorema. *Konvergavimo greičio optimalus tolygusis įvertis yra toks*:

$$\forall (x,n) \in \mathbf{A}\colon\ -\frac{4}{n} < \Delta_n(x) < \frac{1}{48n}.$$

Šiame įvertyje konstantos yra nepagerinamos.

Į klausimą, ar egzistuoja optimalus netolygusis įvertis (žr. [2]) atsako 2 teorema.

2 teorema. *Funkcijos $\Delta_n(x) - \frac{L_1(x)}{n}$ optimali eilė yra $\frac{1}{n^2}$, tačiau neegzistuoja tokia baigtinė konstanta c, kad $\forall (x,n) \in \mathbf{A}$: $|\Delta_n(x) - \frac{L_1(x)}{n}| < \frac{c}{n^2}$, t.y., neegzistuoja optimalus netolygusis konvergavimo greičio įvertis. Bet egzistuoja optimalus netolygusis konvergavimo greičio įvertis iš dešinės*:

$$\forall (x,n) \in \mathbf{A}\colon\ \Delta_n(x) \leqslant \frac{L_1(x)}{n}. \tag{*}$$

Įrodymas. Skaičiuojame ribą

$$L_2(x) := \lim_{n\to\infty} n^2 \cdot \left(\Delta_n(x) - \frac{L_1(x)}{n}\right) = \frac{x(19x^4 - 105x^3 + 50x^2 - 60x + 15)}{(1-x)^4}.$$

Funkcija $L_2(x)$ nėra tapatingai lygi nuliui ir yra baigtinė $\forall x \in \mathbf{U}$, todėl funkcijos $\Delta_n(x) - \frac{L_1(x)}{n}$ eilė yra $\frac{1}{n^2}$. Nelygybė (*) gaunama iš $n \cdot \Delta_n(x)$ monotoniškumo pagal n ir lygybės $\lim_{n\to+\infty} n \cdot \Delta_n(x) = L_1(x)$.

Išvada. Remiantis 1 ir 2 teoremomis gauname, kad:

$$\Delta_n(x) = \frac{L_1(x)}{n} + \frac{L_2(x)}{n^2} + \mathrm{o}\left(\frac{1}{n^2}\right),$$
$$-\frac{4}{n} < \Delta_n(x) \leqslant \frac{L_1(x)}{n}, \quad -\frac{n}{2} < x < 0, \ n \geqslant 2.$$

Literatūra

[1] H.M. Barakat, E.M. Nigm, Convergence of random extremal quotient and product, *Journal of Statistical Planning and Inference*, **81**, 209–221 (1999).
`http://fmf.ktu.lt/vilkas/Straipsniai/Barakat_Nigm.pdf`

[2] R. Vilkas, A. Aksomaitis, Konvergavimo greičio įverčio radimo procedūra, *Liet. Mat. Rink.*, **41** (spec. nr.), 485–490 (2001).
`http://fmf.ktu.lt/vilkas/Straipsniai/Vilkas_Aksomaitis_2001.pdf`

[3] R. Vilkas, Atsitiktinių maksimumų sandaugos konvergavimo greičio įvertis, in: *Matematika ir matematinis modeliavimas-2001*, konferencijos pranešimų medžiaga, Kaunas (2001), pp. 31–33.
`http://fmf.ktu.lt/vilkas/Straipsniai/Konf_2001_KTU.doc`

Optimal estimate of the convergence rate of statistic $\rho_N = |Z_N W_N|$

R. Vilkas, A. Aksomaitis

The optimal estimate of the convergence rate of distribution function of statistic ρ_N is given (for one of the examples in [1]) and unimprovable constants were obtained. Besides, the optimal non-uniform estimate from the right was found together with the unimprovable constant.