Liet. matem. rink, **46**, *spec. nr.*, 2006, 263–266

Didžiausių narių imtyje santykio pasiskirstymas ir momentų savybės

Kęstutis GADEIKIS (VU)
el. paštas: gadeikis@ldr.lt

Reziumė. Šiame straipsnyje nagrinėjamas prof. V. Paulausko ir bendraautorių ([1], [2]) pasiūlytas naujas įvertis pasiskirstymo „uodegos“ indeksui nustatyti, suformuluojamos bei įrodomos didžiausių narių grupelėje santykio k-tųjų momentų savybės, kai $k \geqslant 0$.

Nagrinėsime pasiskirstymo funkcijų F, didelėms argumento x reikšmėms tenkinančių sąryšį

$$\bar{F}(x) \overset{def}{=} 1 - F(x) = x^{-\alpha} L(x),$$

klasę. Koeficientas $\alpha > 0$ yra vadinamas „uodegos“ indeksu, o funkcija L begalybės aplinkoje yra lėtai kintanti:

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1 \quad \forall t > 0.$$

Vienas iš paprasčiausių šios klasės skirstinių – Pareto, kurio pasiskirstymo funkcijos „uodega“ $\bar{F}(x) = (\frac{c}{x})^\alpha$, $c > 0$, $x > c$. Nesunku pastebėti, kad „uodegos“ indeksui α neviršijant dviejų atsitiktiniai dydžiai neturi dispersijos, o vieneto – ir vidurkio.

Nemažo dėmesio sunkia „uodega“ pasižymintys atsitiktiniai dydžiai susilaukė, kai buvo pastebėta, jog akcijų grąžos svyravimai vertybinių popierių biržose, vandens lygio svyravimai, vėjo greitis, ozono ar anglies monoksido koncentracija ore, o taip pat katastrofinio pobūdžio reiškiniai pasižymi laipsniniu greičiu gęstančiomis pasiskirstymo funkcijomis. Tai reiškia, jog ekstremalios reikšmės pasitaiko gerokai dažniau negu nagrinėjant atsitiktinius dydžius, kurių pasiskirstymo funkcijos gęsta eksponentiškai.

Tarkime, kad $X_1, X_2, \ldots, X_N$ yra nepriklausomi vienodai pasiskirstę atsitiktiniai dydžiai, o $X_{N,1} \leqslant X_{N,2} \leqslant \ldots \leqslant X_{N,N}$ – jų variacinė seka. Vienas iš labiausiai žinomų ir dažniausiai praktikoje taikomų parametro $\gamma = 1/\alpha$ Hilo įvertinys atrodo taip:

$$\gamma_{N,k} = \frac{1}{k-1} \sum_{i=1}^{k-1} \ln X_{N,N-i+1} - \ln X_{N,N-k+1}.$$

V. Paulausko ir bendraautorių pasiūlytas įvertis (žr. [1], [2]) iš esmės skiriasi nuo kitų, dažniausiai užrašomų naudojant logaritmines išraiškas, „uodegos“ indekso įverčių. Padalinkime atsitiktinių dydžių seką $X_1, \ldots, X_N$ į n lygių grupių $V_1, \ldots, V_n$, turinčių po m atsitiktinių dydžių, paprastumo dėlei laikydami, jog $N = nm$. Tegu

$$M_{ni}^{(1)} = \max\{X_j\colon X_j \in V_i\},$$

o $M_{ni}^{(2)}$ pažymėkime antrąjį pagal dydį tos pačios grupės V_i elementą. Tada apibrėžkime

$$\kappa_{ni} = \frac{M_{ni}^{(2)}}{M_{ni}^{(1)}}, \quad S_n = \sum_{i=1}^{n} \kappa_{ni}, \quad Z_n = n^{-1} S_n.$$

Iš [2] straipsnyje pateikiamo įrodymo seka, jog kai $N \to \infty$, $m \to \infty$ ir n auga į begalybę kaip, pavyzdžiui, laipsninė N funkcija ($n = CN^b$, $C > 0$, $b \in (0; 1)$), tada

$$Z_n \xrightarrow{b.t.} \frac{\alpha}{\alpha + 1}. \tag{1}$$

Pastarajame straipsnyje taip pat pateikiamas atsitiktinių dydžių κ_{ni}, $1 \leqslant i \leqslant n$, konvergavimas pagal pasiskirstymą, gaunamas pasiremiant [3] straipsnio lema apie variacinės sekos narių pasiskirstymą:

$$\kappa_{ni} \xrightarrow{d} \Big(\frac{\lambda_1}{\lambda_1 + \lambda_2}\Big)^{1/\alpha}, \tag{2}$$

kai $m \to \infty$. Čia λ_1 ir λ_2 yra nepriklausomi eksponentiniai atsitiktiniai dydžiai, kurių vidurkiai lygūs vienetui. Parodysime, jog Pareto atsitiktinių dydžių atveju viskas yra gerokai paprasčiau – konvergavimą pakeičia lygybės ženklas. Šis rezultatas yra žinomas ir naudojamas kaip pagalbinis antrajam teiginiui, kuris jau formuluojamas bei įrodomas grynai prof. V. Paulausko ir bendraautorių pasiūlyto naujojo „uodegos“ indekso įverčio savybėms pagrįsti. Neradus ką pacituoti, žemiau pateikiamas elementarus pirmojo teiginio įrodymas.

1 TEIGINYS. *Tarkime, jog turime m nepriklausomų Pareto atsitiktinių dydžių $X_1, X_2, \ldots, X_m$, kurių pasiskirstymo funkcija $F(x) = 1 - x^{-\alpha}$, $\alpha > 0$, $x \geqslant 1$. Pažymėkime didžiausią atsitiktinį dydį $X_1^* = \max\{X_1, X_2, \ldots, X_m\}$, antrą pagal didumą atsitiktinį dydį pažymėkime X_2^* ir apibrėžkime atsitiktinį dydį*

$$\kappa = \frac{X_2^*}{X_1^*}.$$

Tada nepriklausomai nuo atsitiktinių dydžių skaičiaus m

$$P(\kappa \leqslant x) = x^{\alpha}, \quad x \in [0; 1], \ \ ir \ \ E\kappa = \frac{\alpha}{\alpha + 1}.$$

Įrodymas. Pažymėkime $X_1^* \geqslant X_2^* \geqslant \ldots \geqslant X_m^*$ nagrinėjamų atsitiktinių dydžių $X_1, X_2, \ldots, X_m$ variacinę seką mažėjimo tvarka. Atsitiktiniai dydžiai $X_1, X_2, \ldots, X_m$

tokią variacinę seką gali sudaryti $m(m-1)\dots 1 = m!$ skirtingais būdais. Kadangi pažymėjimų sukeitimas skaičiuojant integralą jo reikšmės nekeičia,

$$
\begin{aligned}
P(\kappa \leqslant x) &= m! \int_1^\infty \int_{x_m}^\infty \int_{x_{m-1}}^\infty \dots \int_{x_3}^\infty \int_{\frac{1}{x}x_2}^\infty \mathrm{d}F_1(x)\,\mathrm{d}F_2(x)\dots \mathrm{d}F_m(x) \\
&= x^\alpha m! \int_1^\infty \int_{x_m}^\infty \int_{x_{m-1}}^\infty \dots \int_{x_3}^\infty \alpha x_2^{-2\alpha}\,\mathrm{d}x_2\,\mathrm{d}F_3(x)\dots \mathrm{d}F_m(x) \\
&= x^\alpha\ \frac{m!}{2} \int_1^\infty \int_{x_m}^\infty \int_{x_{m-1}}^\infty \dots \int_{x_4}^\infty \alpha x_3^{-3\alpha}\,\mathrm{d}x_3\,\mathrm{d}F_4(x)\dots \mathrm{d}F_m(x) \\
&= \dots = x^\alpha\ \frac{m!\cdot 1}{2\cdot 3\cdot \ldots \cdot m} = x^\alpha,
\end{aligned}
$$

čia $F_1, F_2, \dots, F_m$ yra atsitiktinių dydžių $X_1, X_2, \dots, X_m$ pasiskirstymo funkcijos. Tada atsitiktinio dydžio κ vidurkis

$$
E\kappa = \int_0^1 \alpha x^\alpha\,\mathrm{d}x = \frac{\alpha}{\alpha+1}\ .
$$

Dabar pastebėkime, jog nesunkiai galime apskaičiuoti ir atsitiktinių dydžių κ momentus:

$$
E\kappa^k = \int_0^1 x^k \alpha x^{\alpha-1}\,\mathrm{d}x = \frac{\alpha}{\alpha+k},
$$

kai $k \geqslant 0$. Tuo pačiu randame atsitiktinio dydžio κ^k dispersiją:

$$
D\kappa^k = E\kappa^{2k} - \left(E\kappa^k\right)^2 = \frac{\alpha}{\alpha+2k} - \left(\frac{\alpha}{\alpha+k}\right)^2 = \frac{\alpha k}{(\alpha+k)^2(\alpha+2k)}.
$$

Toliau nagrinėsime bendrąjį atvejį, kai atsitiktiniai dydžiai nebūtinai yra Pareto. Prisimindami konvergavimo pagal pasiskirstymą (2) faktą, tęskime skaičiavimus toliau:

$$
\begin{aligned}
P\Big(\frac{\lambda_1}{\lambda_1+\lambda_2} \leqslant x\Big) &= \int_0^\infty \int_{\frac{x_1(1-x)}{x}}^\infty \mathrm{e}^{-x_1}\mathrm{e}^{-x_2}\,\mathrm{d}x_2\,\mathrm{d}x_1 \\
&= \int_0^\infty \mathrm{e}^{-x_1}\mathrm{e}^{-\frac{x_1(1-x)}{x}}\,\mathrm{d}x_1 = \int_0^\infty \mathrm{e}^{-\frac{x_1}{x}}\,\mathrm{d}x_1 = x,
\end{aligned}
$$

kai $0 < x \leqslant 1$. Tada pastebėkime, jog

$$
P\bigg(\Big(\frac{\lambda_1}{\lambda_1+\lambda_2}\Big)^{1/\alpha} \leqslant x\bigg) = P\Big(\frac{\lambda_1}{\lambda_1+\lambda_2} \leqslant x^\alpha\Big) = x^\alpha.
$$

Kadangi su šiuo pasiskirstymu jau susidūrėme, belieka tinkamai suformuluoti ką tik įrodytą teiginį.

2 TEIGINYS. *Tarkime, jog turime m nepriklausomų teigiamų atsitiktinių dydžių $X_1, X_2, \ldots, X_m$, kurių pasiskirstymo „uodegos“ indeksas yra lygus α. Pažymėkime didžiausią atsitiktinį dydį $X_1^* = \max\{X_1, X_2, \ldots, X_m\}$, antrą pagal didumą atsitiktinį dydį pažymėkime X_2^* ir apibrėžkime atsitiktinį dydį*

$$\kappa = \frac{X_2^*}{X_1^*}.$$

Tada atsitiktinis dydis κ, kai $m \to \infty$, pagal pasiskirstymą konverguoja į atsitiktinį dydį χ, kuriam yra teisinga

$$P(\chi \leqslant x) = x^\alpha \ \ ir \ \ E\chi^k = \frac{\alpha}{\alpha + k}, \quad k \geqslant 0.$$

Remiantis pirmuoju teiginiu dabar galima paaiškinti, kodėl nustatant Pareto atsitiktinių dydžių indekso reikšmę narių grupelėje skaičių m geriausia imti lygų dviem. Kaip matyti iš įrodytų rezultatų, didėjant narių skaičiui grupelėje atsitiktinių dydžių κ pasiskirstymas išlieka tas pats, tačiau mažėja pačių grupelių skaičius. Skaičiuojant vidutinę atsitiktinių dydžių κ reikšmę Z_n, akivaizdu, jog rekomenduojamu atveju $m = 2$ turėsime mažiausią „uodegos“ indekso įverčio dispersiją, tuo tarpu vidurkis išliks toks pat.

Literatūra

1. V. Paulauskas, A new estimator for tail index, *Acta Appl. Math.*, **79**, 55–67 (2003).
2. Yu. Davydov, V. Paulauskas, A. Račkauskas, More on p-stable convex sets in Banach spaces, *J. of Theor. Probability*, **13**, 39–64 (2001).
3. R. LePage, M. Woodroofe, J. Zinn, Convergence to a stable distribution via order statistics, *Ann. Probab.*, **9**(4), 624–632 (1981).

SUMMARY

K. Gadeikis. Distribution and moment characteristics of a quotient of heavy-tailed random variables

Let us deal with positive i.i.d. random variables $X_1, \ldots, X_m$ having a tail index α. Calculating a quotient of two biggest members in the set, we obtain a new random variable and investigate its distribution and moment properties. The method itself was proposed by prof. V.Paulauskas and co-authors in [1] and [2].

Keywords: tail index, heavy tail, Pareto.