Liet. matem. rink., **44**, *spec. nr.*, 2004, 584–588

Imčių iš baigtinių visumų statistikų kvazigardelinių skirstinių analizė

Jurgita TURKUVIENĖ, Algimantas BIKELIS (VU)
el. paštas: jurgutet@takas.lt

1. Trumpa apžvalga

Kadangi daugelio statistikų skirstinių negalima užrašyti pakankamai aiškia forma, tai tenka naudoti įvairius asimptotinius skleidinius, gaunamus, naudojant ortogonalių daugianarių sistemas, pavyzdžiui, Edgeworth skleidinius. Kaip žinoma iš [1–10], Edgeworth skleidiniai naudojami kvantilių aproksimacijai (Cornish–Fisher tipo skleidiniai), parametrų vertinimui, pasikliautinų intervalų konstravimui ir hipotezių tikrinimui. Norint išspręsti šiuos uždavinius, reikia sukonstruoti „ilgus" asimptotinius skleidinius, pavyzdžiui, iš 5–6 narių. Iš bendrosios teorijos [11] žinoma, kad, konstruojant tokius asimptotinius skleidinius, būtina atsižvelgti į stebimų atsitiktinių dydžių įgyjamų reikšmių savybes, pavyzdžiui, gardeliškumą. [12], 237 psl., autoriai pateikė pavyzdį, rodantį, kad, konstruojant „ilgus" Edgeworth skleidinius, reikia atsižvelgti į narių kvazigardeliškumą (taip pat žiūr. [13]).

2. Apvertimo formulė

Mes nagrinėjame atsitiktines imtis iš baigtinių visumų serijų:

$$\vec{a}_{N_\nu} = (a_{\nu 1}, a_{\nu 2}, \ldots, a_{\nu N_\nu}), \qquad \nu = 1, 2, \ldots.$$

N_ν – visumos elementų skaičius. Imtis iš baigtinių visumų serijos pirmasis nagrinėjo J. Hájek [14]. Mes pratęsime jo svarstymus ir pastebėsime, kad kiekviena baigtinė realiųjų skaičių aibė

$$a_{\nu 1}, a_{\nu 2}, \ldots, a_{\nu N_\nu}$$

turi sveikąją tiesiškai nepriklausomą bazę [15, 30 psl.], t.y., ją galima užrašyti kaip

$$a_{\nu j} = (\vec{b}_\nu, \vec{e}) + (\vec{\beta}_\nu, \vec{m}_{\nu j}), \qquad j = 1, 2, \ldots, N_\nu. \tag{1}$$

$(\vec{b}_\nu, \vec{e})$ – skaliarinė sandauga, $\vec{b}_\nu$, $\vec{e}$, $\vec{\beta}_\nu$, $\vec{m}_{\nu j}$ – k-mačiai vektoriai erdvėje R^k. Be to $\vec{m}_{\nu j}$ koordinatės įgyja tik sveikas reikšmes $0, \pm 1, \pm 2, \ldots$; $\vec{e} = (1, \ldots, 1) \in R^k$. $\vec{\beta}_\nu$ – bazė, t.y., $(\vec{\beta}_\nu, \vec{m}_{\nu j}) = 0$ tada ir tik tada, kai $\vec{m}_{\nu j} = \vec{0} = (0, \ldots, 0) \in R^k$. Jei $\vec{b}_\nu$ ir $\vec{\beta}_\nu$ duoti, tai (1) išraiška yra vienintelė. Toliau baigtinės visumos elementus žymėsime $a_1, a_2, \ldots, a_N$, t.y., $a_{\nu j} = a_j$.

Nors mokslinėje literatūroje, praktikoje yra nagrinėjamos įvairios imtys, tačiau mes apsiribosime tik dviejomis J. Hájek nagrinėtomis imtimis: paprastąja atsitiktine n tūrio imtimi ir atsitiktinio tūrio Poisson imtimi. Pažymėkime atsitiktinį indikatorių vektorių $\vec{I} = (I_1, I_2, \ldots, I_N)$, t.y., $P\{I_j = 1\} = n/N$, $P\{I_j = 0\} = 1 - n/N$ ir $P\{\vec{I} = \vec{i}\} = 1/C_N^n$, kur vektorius $\vec{i}$ turi n koordinačių lygių vienetui, o likusios lygios nuliui. Mus domina tokios sąlyginės (paprastosioms atsitiktinėms imtims) ir besąlyginės (Poisson imtims) pasiskirstymo funkcijos:

$$P\big\{(\vec{a}, \vec{I}) < x \mid (\vec{I}, \vec{e}) = n\big\}$$

ir

$$P\big\{(\vec{a}, \vec{I}) < x\big\}.$$

$\vec{e} = (1, \ldots, 1) \in R^N$ ir $(\vec{a}, \vec{I})$ – vektorių $\vec{a}$ ir $\vec{I}$ iš N-matės Euklido erdvės R^N skaliarinė sandauga. Naudodami (1), mes gauname:

$$P\big\{(\vec{a}, \vec{I}) < x \mid (\vec{I}, \vec{e}) = n\big\} = \sum_{(\vec{a},\vec{l})<x,\,(\vec{l},\vec{e})=n} \frac{P\{\vec{I} = \vec{l}\}}{P\{(\vec{I}, \vec{e}) = n\}},$$

ir

$$\begin{aligned} &P\big\{(\vec{a}, \vec{I}) = c \mid (\vec{I}, \vec{e}) = n\big\} \\ &= \frac{1}{P\{(\vec{I}, \vec{e}) = n\}} \frac{\beta_1 \ldots \beta_k}{(2\pi)^{k+1}} \int_{-\pi}^{\pi} \int_{\frac{-\pi}{\beta_1}}^{\frac{\pi}{\beta_1}} \cdots \int_{\frac{-\pi}{\beta_k}}^{\frac{\pi}{\beta_k}} \mathrm{e}^{-i\theta n - in(\vec{b}\vec{t}, \vec{e}) - i(\vec{\beta}\vec{t}, \vec{m})} \\ &\quad \times \prod_{j=1}^{N} \big(q + p\mathrm{e}^{i\theta + i(\vec{b}\vec{t}, \vec{e}) + i(\vec{\beta}\vec{t}, \vec{m}_j)}\big)\, \mathrm{d}\theta\, \mathrm{d}t_1 \ldots \mathrm{d}t_k, \end{aligned} \tag{2}$$

kur $c = n(\vec{b}, \vec{e}) + (\vec{\beta}, \vec{m})$, $\vec{t} = (t_1, t_2, \ldots, t_k)$, $\vec{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)$ ir $\vec{\beta}\vec{t} = (\beta_1 t_1, \beta_2 t_2, \ldots, \beta_k t_k)$.

Analogiškai gauname

$$\begin{aligned} P\{(\vec{a}, \vec{I}) = c\} &= \frac{\beta_1 \ldots \beta_k}{(2\pi)^k} \int_{\frac{-\pi}{\beta_1}}^{\frac{\pi}{\beta_1}} \cdots \int_{\frac{-\pi}{\beta_k}}^{\frac{\pi}{\beta_k}} \mathrm{e}^{-in(\vec{b}\vec{t}, \vec{e}) - i(\vec{\beta}\vec{t}, \vec{m}))} \\ &\quad \times \prod_{j=1}^{N} \big(q + p\mathrm{e}^{i(\vec{b}\vec{t}, \vec{e}) + i(\vec{\beta}\vec{t}, \vec{m}_j)}\big)\, \mathrm{d}t_1 \ldots \mathrm{d}t_k, \end{aligned} \tag{3}$$

(2) ir (3) lygybės naudojamos „ilgų" asimptotinių skleidinių konstravimui. J. Hájek [14] parodė, kad, esant tam tikroms sąlygoms, pasiskirstymo funkcijų $P\{(\vec{a}, \vec{I}) < x \mid (\vec{I}, \vec{e}) = n\}$ ir $P\{(\vec{a}, \vec{I}) < x\}$ ribos sutampa, kai $n \to \infty$ ir $N - n \to \infty$. Mes sukonstravome skirtumo $P\{(\vec{a}, \vec{I}) \in B \mid (\vec{I}, \vec{e}) = n\} - P\{(\vec{a}, \vec{I}) \in B\}$ asimptotinį skleidinį, kur B – Borelio aibė erdvėje R^1.

3. Statistikų skirstinių sąsūkos

Kadangi

$$P\{(\vec{a},\vec{I})\in B\mid(\vec{I},\vec{e})=n\}=\frac{P\{(\vec{a},\vec{I})\in B,(\vec{I},\vec{e})=n\}}{P\{(\vec{I},\vec{e})=n\}}$$

$$=\frac{1}{P\{(\vec{I},\vec{e})=n\}}\int\int_{B\times\{n\}}\mathrm{d}\mathop{\boxed{\ast}}_{j=1}^{N}F_j(u,v),$$

kur $F_j(u,v)=P\{a_jI_j<u,I_j<v\}$ – pasiskirstymo funkcija su charakteringąja funkcija $q+pe^{ita_j+i\theta}$ ir $*$ – pasiskirstymo fukcijų $F_j(u,v)$ sąsūkos ženklas, tai pakanka sukonstruoti asimptotinį sąsūkos $\boxed{\ast}_{j=1}^{N}F_j(u,v)$ skleidinį. Aproksimacijai naudosime sąsūką $\boxed{\ast}_{j=1}^{N}(G_j(u)\cdot Q(v))$, kur $G_j(u)=P\{a_jI_j<u\}$ ir $Q(v)=P\{I_j<v\}$. I_j – atsitiktinis Bernulio dydis su tikimybe $P\{I_j=1\}=n/N$. Tarkime, kad $n<N/2$. Jei ši sąlyga patenkinta, tai egzistuoja apibendrinti matai $G_j^{-1}(u)$ ir $Q^{-1}(v)$ tokie, kad $G_j*G_j^{-1}(u)=E(u)$ ir $Q*Q^{-1}(v)=E(v)$, kur $E(u)$ ir $E(v)$ – pasiskirstymo funkcijos, išsigimusios taške nulis.

4. Skleidiniai Appell daugianariais

Skirtumo $P\{(\vec{a},\vec{I})\in B\mid(\vec{I},\vec{e})=n\}-P\{(\vec{a},\vec{I})\in B\}$ asimptotiniam skleidiniui konstruoti panaudosime Appell daugianarius (smulkiau žiūr. [13]). Tegu $P_1,P_2,\dots,P_n$ – pasiskirstymo funkcijos, $G_1,G_2,\dots,G_n$ – apibendrinti matai tokie, kad $G_j*G_j^{-1}=E$;

$$Q_n=\sum_{j=1}^{s}\frac{(-1)^{j+1}}{j}\sum_{l=1}^{n}\left(\sqrt[s]{n}(P_l-G_l)*G_l^{-1}\right)^{*j}\left(\frac{1}{\sqrt[s]{n}}\right)^j;$$

$$q_{jk}(s)=\sum_{\nu_1+\ldots+\nu_j=k,\nu_j=0,1,\ldots}d_j(\nu_1,\nu_2,\dots,\nu_j),$$

$$d_j(\nu_1,\nu_2,\dots,\nu_j)=\frac{(\frac{1}{s+1})^{\nu_1}}{\nu_1!}\cdots\frac{(\frac{1}{s+1})^{\nu_j}}{\nu_j!};$$

$$P^{*n}=\mathop{\boxed{\ast}}_{j=1}^{n}P_j,\qquad G^{*n}=\mathop{\boxed{\ast}}_{j=1}^{n}G_j;$$

$$\rho=\sup_{1\leqslant l\leqslant n}Var\left|\sqrt[s]{n}(P_l-G_l)*G_l^{-1}\right|.$$

4.1 teorema. *Tegu* $\rho n^{\frac{1}{s}}<1$, *tada*

$$P^{*n}(B)=G^{*n}*\mathrm{e}^{Q_n}*\left\{E+\sum_{j=1}^{\infty}\left(\frac{1}{\sqrt[s]{n}}\right)^j(-1)^{j+s+1}\sum_{k=1}^{j}d_j(\nu_1,\nu_2,\dots,\nu_j)\right.$$

$$\times \prod_{\alpha=1}^{j} \left[\frac{1}{n} \sum_{l=1}^{n} (\sqrt[s]{n}(P_l - G_l) * G_l^{-1})^{s+\alpha} \right]^{*\nu_\alpha} \Bigg\} (B)$$

visiems $B \in \boldsymbol{B}(R^k)$, $s = 1, 2, 3, \ldots$. *Pirmasis asimptotinio skleidinio narys yra*

$$U_n^{(1)} = \frac{(-1)^s}{s+1} \frac{1}{\sqrt[s]{n}} \left[\frac{1}{n} \sum_{l=1}^{n} \left(\sqrt[s]{n}(P_l - G_l) * G_l^{-1} \right)^{s+1} \right] * G^{*n} * \mathrm{e}^{Q_n}(B).$$

4.2 teorema. *Tegu* $\sqrt[s]{n}\rho < 1$ *ir* $n \geqslant 1$, *tada*

$$\sup_{\vec{x}} \left| \left(P^{*n} - G^{*n} * \mathrm{e}^{Q_n} - U_n^{(1)} \right)(A - \vec{x}) \right| \leqslant \left[\frac{\rho}{1-\rho} + \frac{\mathrm{e}^{-\frac{1}{13}}}{\sqrt{2\pi}} \cdot \frac{s+1}{(1-\rho)^2} \cdot \frac{\rho^s n^{\frac{s-1}{s}}}{(1 - n\rho^{s+1})^2} \right]$$

$$\times \frac{1}{\sqrt[s]{n}} \left\{ \frac{1}{n} \sum_{l=1}^{n} \sup_{\vec{x}} \left| (\sqrt[s]{n}\left(((P_l - G_l) * G_l^{-1})^{s+1} * G^{*n} * \mathrm{e}^{Q_n} \right)(A - \vec{x}) \right| \right\}$$

visoms Borelio aibėms $A \subset R^k$, $s = 1, 2, 3, \ldots$.

Akivaizdu, kad šios teoremos leidžia naudoti tą faktą, kurį panaudojo ir J. Hájek [14] – pereiti nuo priklausomų prie nepriklausomų atsitiktinių vektorių sumų su pasiskirstymo funkcija

$$\boxed{\prod_{j=1}^{N}}^{*} \left(G_j(u) \cdot Q(v) \right).$$

Pastebėsime, kad dvimatis skirstinys yra kvazigardelinis, ir jo analizei naudojame (2) ir (3) apvertimo formules. Naudodami C. Esseen idėjas [11], iš lokalinių ribinių teoremų gauname integralines ribines teoremas. Toliau įrodinėjant, būtina panaudoti Euler–Maclourin sumavimo formules (žiūr. [8]). Taip gauname „ilgus“ asimptotinius Edgeworth skleidinius sumų iš baigtinių visumų pasiskirstymo funkcijoms.

Literatūra

1. E.A. Cornish, R.A. Fisher, Moments and cumulants in the specification of distributions, *Revue de l'Institute Internal. de Statist.*, **4**, 307–320 (1937).
2. М.Г. Кендалл, А. Стюарт, *Теория распределения*, Наука, М. 588 (1966).
3. Л. Х. Большев, О преобразованиях случайных величин, *Теор. верю и ее прим.* **4**(2), 136–149 (1959).
4. E.V. McCune, H.L. Gray, Cornish–Fisher and Edgeworth expansions, in: *Encyclopedia of Statistics Sciences*, II, (1983), pp. 188–193.
5. J. Pfanzagl, Asymptotic expansions in parametric statistical theory, *Develop. Statist*, **3**, 1–97 (1980).
6. R.N. Bhattacharya, Some recent results on Cramer–Edgeworth expansions with applications, in: P.R. Krishnaiah (Ed.), *Multivariate Analysis* VI, Amserdam, North-Holland (1985), pp. 57–75.
7. R.N. Bhattacharya, M. Denker, *Asymptotic Statistics*, Birkhauser, Boston (1990).

8. R.N. Bhattacharya, R. Ranga Rao, *Normal Approximation and Asymptotic Expansions*, Krieger Malaber, Florida (1986).
9. В. Калленберг, Интерпретация и использование разложений Энжворта, *Теор. вер. и ее прим.*, **37**(1), 179–181 (1992).
10. П. Ембрехтс, К. Клюппелберг, Некоторые аспекты страховой математики, *Теор. вер. и ее прим.*, **38**(2), 374–416 (1993).
11. C.-G. Esseen, Fourier analysis of distribution functions: a mathematical study of the Laplace–Gaussian law, *Acta Math.*, 1–125 (1945).
12. Б.В. Гнеденко, А.Н. Колмогоров, *Предельные распределения для сумм независимых случайных величин*, Гостехиздат, Москва 264 (1949).
13. A. Bikelis, Asymptotic expansions for distribution of statistics, in: *Proceedings of the XXXVI Conference of Lithuanian Mathematicians*, 1–24 (1995).
14. J. Hājek, Limiting distributions in simple random sampling for a finite population, *Magyar Tud Akad Mat Kutato Int Korl*, **5**, 361–374 (1960).
15. Б.М. Левитан, *Почти периодические функции и дифференциальные уравнения*, МГУ, Москва (1978).

SUMMARY

J. Turkuvienė, A. Bikelis. Analysis of quasi-lattice distributions of statistics from finite population

Edgeworth expansions are used for approximation of quantiles, estimation of parameters, construction of confidence intervals and testing hypothesis. Paper shows how to construct "long" Edgewort asymptotic expansions.

Keywords: simple random sample, Poisson sample, Edgeworth expansions, Cornish–Fisher expansions, quasi-lattice distributions, Appell polynomials.