Liet. matem. rink., T. 41, spec. nr., 519–525

Apie perėjimo tikimybių pasikliautinius intervalus mirties priežasčių sindrominėje analizėje

Algimantas BIKELIS (VU), Sigitas DAPKŪNAS (VU),
Mečislovas MEILŪNAS (VU, VGTU), Danutė STOŠKUVIENĖ (VU)
el. paštas: mecislovas.meilunas@fm.vtu.lt, sigitas.dapkunas@sc.vu.lt

1. Mirties priežasčių sindrominės analizės matematinis modelis

Nagrinėsime žmogaus organizmo biologinių funkcijų degradavimo procesus pasirinktame ligonių kontingente. Tokie procesai apibrėžiami *tanatogeneziniais sindromais* (TGS). *Tanatogenezė* – mirties proceso eiga.

Laikoma, kad liga tampa nepagydoma, kai jos komplikacijos virsta tanatogeneziniais sindromais (TGS). Tanatogenezės pradžioje pažeidžiamos gyvybiškai svarbios – kvėpavimo, kraujotakos ir nervinės reguliacijos funkcijos. Tai pagrindiniai TGS. Kai negrįžtamai pažeidžiami organai, gyvybės procesai nyksta, prasideda terminaliniai TGS. Galop nutrūksta viena šių trijų (kartais kartu dvi ar visos trys) gyvybės funkcijos – kvėpavimo, kraujotakos ir centrinio nervinio reguliavimo, atitinkamai gyvybė užsibaigia kvėpavimo, kraujotakos, centrinės nervinės reguliacijos mirties mechanizmais [1].

Žmogaus gyvybinės funkcijos gali degraduoti dėl genetiškai užprogramuotų amžiaus veiksnių, dėl smurtinio gyvybės procesų pažeidimo ir dėl ligų. Daugumos (apie 90%) žmonių gyvybinių funkcijų pažeidžiama ligų. Žmogaus biologinė degradacija gali nesivystyti, kai liga pagydoma, ir progresuoti iki gyvybės funkcijų nutrūkimo, kai jis miršta. Žmogaus biologinės degradacijos dėl ligų modelis pirmiausia remiasi mirusiųjų ligoninėje ligonių gyvybinių funkcijų pažeidimo ir nutrūkimo eiga. Todėl visos toliau naudojamos sąvokos apibrėžiamos mirusiųjų kontingentui.

Pagrindinė mirties priežasčių sindrominės analizės ir ja besiremiančio biologinės degradacijos modelio prielaida yra ta, kad kiekvienam mirusiąjam konkrečiame tanatogeneziniame lygmenyje galima vienareikšmiškai priskirti apibrėžtą būseną.

Tanatogenezinis procesas suskirstytinas šiais atskirais etapais (lygmenimis):

A – sveikieji žmonės;

Λ – nustatyta liga susirgę ligoniai;

Π – ligoniai, praėję pagrindinį TGS;

T – ligoniai, praėję terminalinį TGS;

M – ligoniai, praėję mirties mechanizmą;

Ω – mirusieji.

Taigi, tarsime, kad kiekvienas ligonis iš nagrinėjamo mirusiųjų kontingento yra praėjęs visus tuos etapus, t.y., ligoniui galima vienareikšmiškai priskirti bet kurio etapo konkrečią būseną. Tokiu būdu, kiekvienas ligoninėje miręs ligonis yra praėjęs per visus lygmenis tokia seka:

$$A \to \Lambda \to \Pi \to T \to M \to \Omega.$$

Tarkime, lygmenyje Λ yra išskirta I būsenų (ligų), Π – J būsenų (pagrindinių sindromų), T – K būsenų (terminalinių sindromų), M – L būsenų (mirties mechanizmų). Laikykime, kad A ir Ω lygmenys neskaidomi.

Tegul atskirose lygmenų būsenose mirusieji buvo pasiskirstę taip:

Liga	Λ_1	Λ_2	...	Λ_I
Ligonių skaičius	n_1^L	n_2^L	...	n_I^L

Pagr. sindr.	Π_1	Π_2	...	Π_J
Ligonių skaičius	n_1^P	n_2^P	...	n_J^P

Termin. sindr.	T_1	T_2	...	T_K
Ligonių skaičius	n_1^T	n_2^T	...	n_K^T

Mirties mechan.	M_1	M_2	...	M_L
Ligonių skaičius	n_1^M	n_2^M	...	n_L^M

Pažymėsime perėjimo iš būsenos Λ_i į Π_j tikimybę $p_{j/i}$, perėjimo iš Π_j į T_k tikimybę $q_{k/j}$, perėjimo iš T_k į M_l tikimybę $r_{l/k}$. Galime sudaryti perėjimo tikimybių matricas

$$P = \begin{bmatrix} p_{1/1} & p_{1/2} & \cdots & p_{1/I} \\ p_{2/1} & p_{2/2} & \cdots & p_{2/I} \\ \cdots & \cdots & \cdots & \cdots \\ p_{J/1} & p_{J/2} & \cdots & p_{J/I} \end{bmatrix}, \qquad Q = \begin{bmatrix} q_{1/1} & q_{1/2} & \cdots & q_{1/J} \\ q_{2/1} & q_{2/2} & \cdots & q_{2/J} \\ \cdots & \cdots & \cdots & \cdots \\ q_{K/1} & q_{K/2} & \cdots & q_{K/J} \end{bmatrix},$$

$$R = \begin{bmatrix} r_{1/1} & r_{1/2} & \cdots & r_{1/K} \\ r_{2/1} & r_{2/2} & \cdots & r_{2/K} \\ \cdots & \cdots & \cdots & \cdots \\ r_{L/1} & r_{L/2} & \cdots & r_{L/K} \end{bmatrix},$$

kurios realizuoja atvaizdavimų grandinę

$$A \xrightarrow{P_0} \Lambda \xrightarrow{P} \Pi \xrightarrow{Q} T \xrightarrow{R} M \xrightarrow{R_0} \Omega.$$

Tokiu būdu, pavyzdžiui, matricos P elementas $p_{j/i}$ yra tikimybė pereiti iš i-tosios ligos į pagrindinį sindromą, kuris yra pažymėtas j-tuoju numeriu. Perėjimas iš A į Λ aprašomas vektoriumi $P_0 = (p_1^0, p_2^0, \ldots, p_I^0)$, perėjimo iš M į Ω – vektoriumi $R_0 = (r_1^0, r_2^0, \ldots, r_L^0)$.

Vaizdumo dėlei galime tarti, kad kiekviena bet kurio lygmens būsena – tai kiaura dėžė, į kurią atsitiktinai gali įkristi ir iš jos iškristi rutuliukas. Iškritęs rutuliukas su tam tikra perėjimo tikimybe gali pakliūti į kurią nors žemesnio lygmens dėžę ir t.t.

Tegul $p_{j/i}$ – perėjimo iš i-tosios būsenos į j-tąją tiksli tikimybės reikšmė, o $\hat{p}_{j/i}$ – jos taškinis įvertis, kurį apibrėžiame kaip santykinį dažnį m_{ij}/m_j, kur m_{ij} – iš i-tosios būsenos atėjusių į j-tąją būseną objektų skaičius, o m_j – bendras j-toje būsenoje esančių objektų skaičius. Laikome, kad perėjimo tikimybės yra pasiskirsčiusios pagal polinominį dėsnį, kurios kiekvienos komponentės marginalinis skirstinys – binominis.

Tegul turime n tūrio imtį. Skaičiuojame tikimybę $p_{k/ij}$ – tikimybę būti k-tojoje būsenoje, praėjus i ir j būsenas $p_{k/ij} = p_{ijk}/p_{ij}$, kur $i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K$. Tada tikimybės įvertis bus

$$\hat{p}_{k/ij} = \frac{m_{ijk}}{m_{ij}}, \quad m_{ij} = \sum_{l=1}^{K} m_{ijl},$$

čia m_{ijk} – individų, esančių būsenoje k praėjus i ir j būsenas skaičius, m_{ij} – individų, esančių būsenoje j ir praėjusių i būseną, skaičius. Į šiuos dažnius galime žiūrėti kaip į atsitiktinius dydžius. Laikome, kad atsitiktinis vektorius M turi polinominį skirstinį su parametrais n, p_{ijk} :

$$\begin{aligned} M &= \{m_{ijk}, i = \overline{1,I},\ j = \overline{1,J},\ k = \overline{1,K}\}. \\ M &\sim M(n; p_{ijk}),\ \sum_{i,j,k} p_{ijk} = 1. \end{aligned}$$

Todėl atsitiktinio dydžio $\hat{p}_{k/ij} = \frac{m_{ijk}}{m_{ij}}$ pasiskirstymo funkcija turi išraišką

$$\begin{aligned} F(x) &= P\{\frac{m_{ijk}}{m_{ij}} \leqslant x\} \\ &= \sum_{\alpha_1=0}^{n} \sum_{\alpha_2=0}^{n} \frac{n!}{\alpha_1!\alpha_2!(n-\alpha_1-\alpha_2)!} p_{ijk}^{\alpha_1} (p_{ij} - p_{ijk})^{\alpha_2} (1 - p_{ij})^{n-\alpha_1-\alpha_2}, \\ & 0 \leqslant \alpha_1 + \alpha_2 \leqslant n, \\ & \alpha_1(1-x) - \alpha_2 x \leqslant 0, \\ & p_{k/ij} = \frac{p_{ijk}}{p_{ij}}, \quad p_{ij} = p_{ijk} + \sum_{l=1, l\neq k}^{K} p_{ijl}. \end{aligned} \tag{1}$$

Pasinaudojus charakteristine funkcija, galima įrodyti, kad

$$M(\hat{p}_{k/ij}) = p_{k/ij} \lfloor 1 - (1 - p_{ij})^n \rfloor.$$

2. Pasikliautinio intervalo radimo algoritmas

Norėdami nustatyti perėjimo tikimybės pasikliautinį intervalą, turime spręsti lygtį

$$F(x) \equiv \sum_{\alpha_1=0}^{n} \sum_{\alpha_2=0}^{n} \frac{n!}{\alpha_1!\alpha_2!(n-\alpha_1-\alpha_2)!} p_{ijk}^{\alpha_1} (p_{ij} - p_{ijk})^{\alpha_2}\ (1 - p_{ij})^{n-\alpha_1-\alpha_2} = \omega, \tag{2}$$

kur $\alpha_j \in N$, $\alpha_1 + \alpha_2 \leqslant n$, $\alpha_1(1-x) - \alpha_2 x \leqslant 0$.

Ją reikia išspręsti x atžvilgiu, kai duotas ω.

Pasižymime $p_{ijk} = p, p_{ji} - p_{ijk} = \Theta$, taip pat apibrėžiame $\lfloor y \rfloor$ („grindys“ – y sveikoji dalis), $\lceil y \rceil$ („lubos“ – y sveikoji dalis + 1) ir lygtį perrašome taip:

$$\sum_{\alpha_1=0}^{\lfloor nx \rfloor} \sum_{\alpha_2=\lceil \frac{1-x}{x}\alpha_1 \rceil}^{\lfloor n-\alpha_1 \rfloor} \frac{n!}{\alpha_1!\alpha_2!(n-\alpha_1-\alpha_2)!} p^{\alpha_1}\Theta^{\alpha_2}(1-p-\Theta)^{n-\alpha_1-\alpha_2} = \omega. \tag{3}$$

Lygties (2) dešinioji pusė gali priklausyti vienam iš dviejų tipų:

1. $\omega = \omega^0$ – lygtis sprendinių neturi, t.y., neegzistuoja x, tenkinantis $f(x) = \omega$.
2. $\omega = \omega^\alpha$ – lygtis turi be galo daug sprendinių.

Tiesės

$$\alpha_2 = \frac{1-x}{x}\alpha_1$$

krypties koeficientą pažymėkime ξ. Taigi, turime $\alpha_2 = \xi\alpha_1$, kur $0 \leqslant \xi < \infty$.

Sudarome Farėjaus (Farey) seką $F_n = \{f_i\}, i = 1, 2, \ldots$ kurią sudaro paprastosios taisyklingos nesuprastinamos trupmenos, kurių vardiklis ne didesnis už n, išdėstytos didėjančia tvarka. Tokios trupmenos atitiks $f(x)$ trūkio taškus. Taip pat sudarome pagalbinę seką

$$\tilde{F}_n = \{\frac{f_i + f_{i+1}}{2}\},$$

kurios elementai yra sekos F_n gretimų elementų aritmetiniai vidurkiai. Taigi, turime

$$f_1 < \tilde{f}_1 < f_2 < \tilde{f}_2 < \ldots < f_i < \tilde{f}_i < f_{i+1} < \ldots$$

(1) lygtį sprendžiame pusiaukirtos tipo metodu.

Priminsime pusiaukirtos algoritmą tolydžios funkcijos atveju.

Jei lygtis $g(x) \equiv f(x) - \omega = 0$ intervalo $[a, b]$ viduje turi šaknį, tai $g(a)g(b) < 0$. Apskaičiuojame $c = \dfrac{a+b}{2}$. Jei $g(c \neq 0)$, tai arba $g(a)g((a+b)/2) < 0$, arba $g((a+b)/2))g(b) < 0$. Renkamės tą intervalą, kuriame ši sąlyga išpildyta ir daliname jį pusiau. Procesą kartojame tol, kol intervalo ilgis nesumažės iki reikiamo tikslumo. Mūsų čia nagrinėjamu atveju $f(x)$ yra trūki funkcija, todėl tam, kad realizuojant pusiaukirtos algoritmą nepataikyti ant trūkio taškų, naudojame tiktai sekos

$$\tilde{F}_n = \{\frac{f_i + f_{i+1}}{2}\}$$

elementus. Tokiu būdu, jeigu $\dfrac{a+b}{2} = f_i$, tai imame $\tilde{f}_i$, o ne f_i, ir taške $\tilde{f}_i$ skaičiuojame funkcijos F reikšmę.

Toliau skirsime du atvejus.

1) Jei smulkindami pradinį intervalą [0,1] priėjome iki intervalo, skiriančio sekos $\tilde{F}_n$ gretimus narius $\lfloor \tilde{f}_i, \tilde{f}_{i+1} \rfloor$ ir turime $F(\tilde{f}_i)F(\tilde{f}_{i+1}) < 0$, tai tvirtiname, kad ieškoma šaknis $x = f_i$.

2) Jei kuriame nors žingsnyje gavome $F(\tilde{f}_i) = \omega$, tai visi intervalo (f_i, f_{i+1}) taškai tenkina (1) lygtį. Tokiu būdu (1) lygties šaknis – intervalinis dydis.

Farėjaus seką sudarome remdamiesi [2]. Jei $\dfrac{\mu}{\gamma}, \dfrac{\mu'}{\gamma'}, \dfrac{\mu''}{\gamma''}$ – sekos F_n gretimi nariai, tai jie tenkina sąryšius

$$\mu'' = \left[(\gamma + n)/\gamma'\right]\mu' - \mu,$$
$$\gamma'' = \left[(\gamma + n)/\gamma'\right]\gamma' - \gamma.$$

Naudodamiesi šiais sąryšiais, paėmę pradines reikšmes $0/1$ ir $1/n$, galime gauti visus sekos F_n narius.

Gali būti taip, kad apskaičiuodami (2) lygties kairioje pusėje esančios sumos dėmenis, turėsime atlikti veiksmus su labai dideliais ir labai mažais skaičiais. Todėl (2') lygtį užrašome, prieš tai pergrupavę jos narius

$$\sum_{\alpha_1=0}^{\lfloor nx \rfloor} \sum_{\alpha_2=\lceil \frac{1-x}{x}\alpha_1 \rceil}^{\lfloor n-\alpha_1 \rfloor} p^{\alpha_1} \frac{\alpha_1+1}{1}\theta \dots \frac{\alpha_1+\alpha_2}{\alpha_2}\theta$$
$$\times \left[\frac{\alpha_1+\alpha_2+1}{1}(1-p-\theta)\dots\frac{n}{n-\alpha_1-\alpha_2}(1-p-\theta)\right] = 0.95.$$

Jos sprendinys – sąlyginė tikimybė įgyti terminalinį sindromą k sergant liga i ir įgijus pagrindinį sindromą j.

3. Rezultatai ir išvados

1, 2 lentelėse pateikti rezultatai gauti, atlikus skaičiavimus imčiai $n = 118$ (1 lentelė), $n = 267$ (2 lentelė).

Lentelių grafos:

1. m_{ij} – kiek mirė ligonių, sirgę liga i, perėję pagrindinį sindromą j.
2. m_{ijk} – kiek mirė ligonių, sirgę liga i, perėję pagrindinį sindromą j ir terminalinį sindromą k.
3. m_{ijk}/n
4. m_{ij}/n
5. $\theta = m_{ij}/n - m_{ij}/n$

6. m_{ijk}/m_{ij}
7. kp1-kp2 – pasikliautinis intervalas, gautas aproksimavus normaliuoju dėsniu [3].
8. kp3-kp4 – pasikliautinis intervalas, gautas pagal [4].
9. kp9-kp10 – suskaičiuotos reikšmės pagal pateiktą algoritmą.

[4] straipsnyje pasiūlyti metodai, kaip, taikant aproksimacijas normaliuoju dėsniu apskaičiuoti perėjimo tikimybių pasikliautinius intervalus. Šiame straipsnyje pateiktas algoritmas, kaip pasikliautinių intervalų galus galima apskaičiuoti tiesiogiai, nesinaudojant aproksimacijomis. Šiuolaikinių kompiuterių galimybės ir paprasčiausi skaitiniai metodai leidžia tai efektyviai atlikti.

1 lentelė. Tikimybės, gautos tikslinant intervalą $[0, 1]$ pasirinkus tikslumą 0.001, $w = 0.925, w = 0.0125, n = 118$.

1	2	3	4	5	6	7		8		9	
3	3	0.0254	0.0254	0.0000	1.0000	0.4385	1.0000	1.0000	1.0000	1.0000	1.0000
4	1	0.0085	0.0339	0.0254	0.2500	0.0455	0.6994	0.1324	0.3676	0	1.0000
4	2	0.0169	0.0339	0.0170	0.5000	0.1500	0.8500	0.3642	0.6358	0	1.0000
88	43	0.3644	0.7458	0.3814	0.4886	0.3869	0.5913	0.4596	0.5176	0.3839	0.5930
88	1	0.0085	0.7458	0.7373	0.0114	0.0020	0.0616	0.0052	0.0175	0	0.3614
88	41	0.3475	0.7458	0.3983	0.4659	0.3653	0.5694	0.4370	0.4948	0.3617	0.5726
88	3	0.0254	0.7458	0.7204	0.0341	0.0117	0.0955	0.0235	0.0446	0	0.0769
10	4	0.0339	0.0847	0.0508	0.4000	0.1682	0.6873	0.3158	0.4841	0.1000	0.7272
10	3	0.0254	0.0847	0.0593	0.3000	0.1078	0.6032	0.2213	0.3787	0	0.6250
10	2	0.0169	0.0847	0.0678	0.2000	0.0566	0.5098	0.1313	0.2687	0	0.5000
10	1	0.0085	0.0847	0.0762	0.1000	0.0178	0.4041	0.0485	0.1515	0	0.3362
4	3	0.0254	0.0339	0.0085	0.7500	0.3006	0.9545	0.6324	0.8676	0	1.0000

2 lentelė. Tikimybės, gautos tikslinant intervalą $[0, 1]$ pasirinkus tikslumą 0.00001, $w = 0.975, w = 0.0125, n = 267$.

1	2	3	4	5	6	7		8		9	
8	1	0.0037	0.0300	0.0263	0.1250	0.0220	0.4710	0.0000	0.3550	0.0000	0.4000
8	4	0.0150	0.0300	0.0150	0.5000	0.2150	0.7850	0.1520	0.8480	0.1250	0.8750
2	2	0.0075	0.0075	0.0000	1.0000	0.3420	1.0000	1.0000	1.0000	1.0000	1.0000
6	4	0.0150	0.0225	0.0075	0.6667	0.3000	0.9030	0.2880	1.0000	0.2000	1.0000
2	2	0.0075	0.0075	0.0000	1.0000	0.3420	1.0000	1.0000	1.0000	1.0000	1.0000
15	13	0.0487	0.0562	0.0075	0.8667	0.6210	0.9630	0.6940	1.0000	0.6679	1.0000
10	7	0.0262	0.0375	0.0113	0.7000	0.3970	0.8920	0.4150	0.9850	0.3750	1.0000
10	1	0.0037	0.0375	0.0338	0.1000	0.0170	0.4040	0.0000	0.2860	0.0000	0.3345
13	3	0.0112	0.0487	0.0375	0.2308	0.0820	0.5030	0.0016	0.4599	0.0000	0.5000
13	4	0.0152	0.0487	0.0337	0.3077	0.1270	0.5760	0.0560	0.5600	0.0719	0.5830
13	2	0.0075	0.0487	0.0412	0.1538	0.0430	0.4220	0.0400	0.3510	0.0000	0.3846
7	3	0.0112	0.0262	0.0150	0.4286	0.1580	0.7500	0.0600	0.7970	0.0000	0.8330
7	1	0.0037	0.0262	0.0225	0.1429	0.0250	0.5130	0.0000	0.4030	0.0000	0.5000
8	2	0.0075	0.0300	0.0225	0.2500	0.0710	0.8910	0.0000	0.5510	0.0000	0.5990
2	2	0.0075	0.0075	0.0000	1.0000	0.3420	1.0000	1.0000	1.0000	1.0000	1.0000
10	1	0.0037	0.0375	0.0338	0.1000	0.0180	0.4040	0.0000	0.2870	0.0000	0.3345
10	5	0.0187	0.0375	0.0188	0.5000	0.2370	0.7630	0.1890	0.8110	0.1670	0.8333
6	6	0.0225	0.0225	0.0000	1.0000	0.6100	1.0000	1.0000	1.0000	1.0000	1.0000
7	1	0.0037	0.0262	0.0225	0.1429	0.0260	0.5130	0.0000	0.4030	0.0000	0.5000
7	2	0.0075	0.0262	0.0187	0.2857	0.0820	0.6410	0.0000	0.6220	0.0000	0.6679
7	1	0.0037	0.0262	0.0225	0.1429	0.0260	0.5130	0.0000	0.4030	0.0000	0.5000
8	5	0.0187	0.0300	0.0113	0.6250	0.3060	0.8630	0.2880	0.9620	0.2500	1.0000
18	18	0.0674	0.0674	0.0000	1.0000	0.8240	1.0000	1.0000	1.0000	1.0000	1.0000
9	8	0.0300	0.0337	0.0037	0.8889	0.5650	0.9800	0.6830	1.0000	0.6250	1.0000
17	16	0.0599	0.0637	0.0038	0.9412	0.7300	0.9900	0.8290	1.0000	0.8000	1.0000

Lentelėse panaudoti realūs duomenys apie ligonius, mirusius Vilniaus universitetinėje Antakalnio ligoninėje pastaraisiais metais.

Už šiame straipsnyje nagrinėto uždavinio biomedicininį formulavimą ir rezultatų aptarimą autoriai nuoširdžiai dėkingi VU profesoriui R. Ptašekui.

Literatūra

[1] R. Ptašekas, *Žmogaus ligų tanatogeneziniai sindromai*, VU leidykla (1998).

[2] R.L. Graham, D.E. Knuth, O. Patashnik, *Concrete Mathematics*, Addison-Wesley (1998).

[3] B.L. Van der Waerden, *Mathematical Statistics*.

[4] V. Bagdonavičius, A. Bikelis, M. Meilūnas, D. Stoškuvienė, On the human's vital functions degradation modeling, *Universite Victor Segalen Bordeaux 2*, Bordeaux (2000).

[5] A. Bikelis, Asymptotic expansions for distributions of statistics, *XXXVI Conference of Lithuanian Mathematical Society*, Vilnius (1995).

On the transition probabilities confidence interval evaluation in the syndromic analysis

A. Bikelis, S. Dapkūnas, M. Meilūnas, D. Stoškuvienė

A confidence interval evaluation problem of transition probabilities is considered.