Liet. matem. rink., T. 40, spec. nr., 482–485

Apibendrinto multiplikatyvaus patikimumo funkcijos modelio taikymas Lietuvos gyventojų išgyvenamumo funkcijos aproksimavimui

Rasa SENKUVIENĖ (LSIC, MII)
el. paštas: rasa.miseikyte@takas.lt

Tarkime, kad individo, gimusio laiko momentu k ir išgyvenusio iki amžiaus t momentinė rizika numirti laiko momentu $k+t$ aprašoma multiplikatyvios rizikos (Cox'o) modeliu:

$$\mu_{\omega k}(t) = \omega \cdot x(k+t)\mu(t),$$

čia $\mu(t) > 0$ – amžiaus funkcija, $x(k+t) > 0$ – funkcija nuo laiko, o ω – individuali konstanta. Tuomet individo išgyvenamumo funkcija bus

$$S_{\omega k}(t) = \exp\left(-\omega \int_0^t x(\tau+k)\mu(\tau)\,\mathrm{d}\tau\right),$$

o vidutinė populiacinė kohortos išgyvenamumo funkcija bus aprašoma individualių išgyvenamumo funkcijų mišiniu:

$$S_k(t) = \int_0^\infty g(\omega)\exp\left(-\omega\int_0^t x(\tau+k)\mu(\tau)\,\mathrm{d}\tau\right)\mathrm{d}\omega,$$

čia $g(\omega)$ – atsitiktinio dydžio ω tankio funkcija. Pažymėkime

$$\Lambda_k(t) = \int_0^t x(\tau+k)\mu(\tau)\,d\tau,$$

sukauptą kumuliatyvią riziką, kai $\omega = 1$. Tuomet

$$G(\Lambda_k(t)) = \mathbf{E}\Big(\exp\big(-\omega\Lambda_k(t)\big)\Big) = S_k(t) \tag{1}$$

bus taisyklinga patikimumo funkcija, taigi aprašys apibendrintą multiplikatyvųjį modelį.

Tarkime, kad $x(t)$ galime pakankamai tiksliai aproksimuoti laiptuota funkcija

$$x(t) = x_j, \quad j \leqslant t < j+1.$$

Tada diskrečiais laiko momentais $t = 0, 1, \ldots$ $\Lambda_k(t)$ galėsime aproksimuoti funkcija

$$\Lambda_k^*(t) = \sum_{j=0}^{t-1} x_{k+j} \int\limits_j^{j+1} \mu(\tau)\,\mathrm{d}\tau = \sum_{j=0}^{t-1} x_{k+j}\lambda_j,$$

čia $\lambda_j = \int_j^{j+1} \mu(\tau)\,\mathrm{d}\tau$.

Aproksimuokime k-tosios kohortos $[t, t+1)$ amžiaus gyventojų mirtingumą $k+t$-ųjų metų $[t, t+1)$ amžiaus gyventojų mirtingumu. Tuomet empirinė k-tosios kohortos tikimybė išgyventi $[t, t+1)$-ąjį laiko momentą bus aproksimuojama dydžiu

$$\hat{p}_{k,t} = 1 - m_{k+t,t}/n_{k+t,t},$$

čia $m_{j,t}$ ir $n_{j,t}$ žymi j-tųjų kalendorinių metų $[t, t+1)$-osios amžiaus grupės individų mirčių skaičių ir gyvenančių intervalo pradžioje skaičių, atitinkamai. Laukiamas $\hat{p}_{k,t}$ vidurkis, gautas iš modelio (1) bus

$$p_{k,t}^* = G\big(\Lambda_k^*(t+1)\big)/G\big(\Lambda_k^*(t)\big). \tag{2}$$

Taikant (2) aproksimaciją vyresnių nei 40 metų amžiaus Lietuvos gyventojų 1980–1996 metų mirtingumo duomenims funkcija $G(\Lambda)$ buvo aproksimuota Hougaar'do [2] pasiūlytu mišiniu:

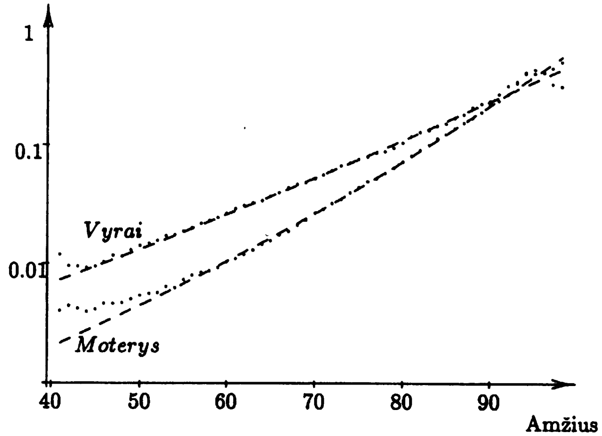
$$G(\Lambda, \theta, \rho^2) = \exp\left[-\frac{1-\theta}{\theta\rho^2}\left(\left(1 + \frac{\rho^2}{1+\theta}\Lambda\right)^{\theta} - 1\right)\right], \tag{3}$$

čia $0 \leqslant \theta \leqslant 1$ ir $\rho^2 \geqslant 0$ – parametrai. Šis mišinys buvo pasirinktas dėl to, kad, turėdamas tik du parametrus gali įgyti labai įvairius pavidalus. Kai $\theta = 0$, šis mišinys virsta Gamma mišiniu $G(t) = [1 + \rho^2\Lambda(t)]^{-\frac{1}{\rho^2}}$, o kai $\theta = 1$ arba $\rho^2 = 0$ išsigimsta į paprastą eksponentinę funkciją $G(t) = \exp[-\Lambda(t)]$. Sukaupta kumuliatyvi rizika, kai $\omega = 1$

$$\Lambda(t) = \int\limits_0^t \mu(\tau)\,\mathrm{d}\tau,$$

diskretiems laiko momentams $t = 0, 1, \ldots$ buvo aproksimuota neparametrine

$$\Lambda(t) = \sum_{j=0}^{t-1} \lambda_j, \tag{4}$$



1 pav. Funkcijos $\log\{-\log(G[\Lambda(t+1)]/G[\Lambda(t)])\}$ reikšmės, kai $G(\Lambda)$ yra (3)-čio, o $\Lambda(t)$ (4)-to bei (5)-to pavidalų (Lietuvos gyventojų 1980–1996 metų mirtingumo aproksimacija).

bei parametrine

$$\Lambda(t) = \exp\left[\alpha + \beta\frac{t^k - t^{-k}}{2k}\right] \tag{5}$$

funkcijomis. Pastaroji aproksimacija buvo pasirinkta, nes buvo viena iš geriausių aproksimuojant kohortinius duomenis [3].

Aproksimuojant 40–98 m. amžiaus Lietuvos gyventojų 1980–1996-ų metų mirtingumo duomenis, iš viso turėjom 903 stebėjimus $\hat{p}_{k,t}$ Buvo minimizuojamas vidutinis kvadratinis nuokrypis

$$\frac{1}{N}\sum_{k,t}\Big(\log\left(-\log(\hat{p}_{k,t})\right)-\log\Big(-\log\left(G(\Lambda_k^*(t+1),\theta,\rho^2)/G(\Lambda_k^*(t),\theta,\rho^2)\right)\Big)\Big)^2, \tag{6}$$

čia N – laisvės laipsniai, gauti, atėmus iš 903 vertinamų parametrų skaičių.

Kadangi empirinė dydžio $\hat{\mu}_{i,t} = \log(-\log(\hat{p}_{i-t,t}))$ priklausomybė nuo amžiaus t beveik tiesinė (žr. 1 pav.), tai galime sukonstruoti empirinį vidutinio kvadratinio nuokrypio (6) analogą. Iš tikro, jei sukonstruosime neparametrinį dydžio $\hat{\mu}_{i,t}$ vidurkio įvertį kaip

$$\mu_{i,t}^* = \left(1 - \frac{2}{\sqrt{6}}\right)\hat{\mu}_{i,t} + \frac{1}{\sqrt{6}}\left(\hat{\mu}_{i,t-1} + \hat{\mu}_{i,t+1}\right)$$

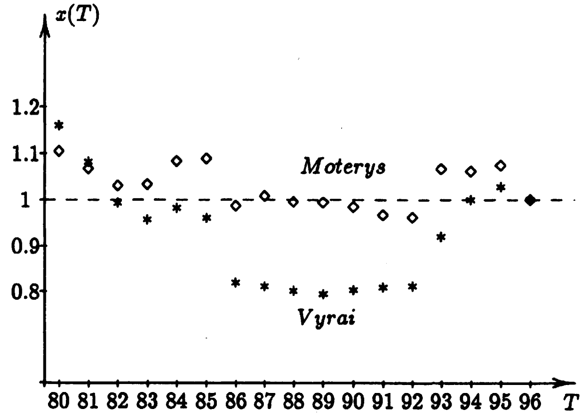
ir tarsime, kad $\hat{\mu}_{i,t}$ stebėjimų paklaidos i-taisiais metais nekoreliuotos pagal amžių, tai $\hat{\mu}_{i,t}^*$ bus beveik nepaslinktas $\hat{\mu}_{i,t}$ įvertis, o

$$\mathbf{E}\Big(\sum_i\sum_{t=t_0}^{T}(\hat{\mu}_{i,t} - \mu_{i,t}^*)^2\Big) =$$

$$= \sum_{i} \left(\sum_{t=t_0}^{T} \sigma_{i,t}^2 + \frac{1}{6} \left(\sigma_{i,T+1}^2 + \sigma_{i,t_0-1}^2 - \sigma_{i,T}^2 - \sigma_{i,t_0}^2 \right) \right),$$

čia $\sigma_{i,t}^2$ yra dydžio $\hat{\mu}_{i,t}$ dispersija, o i_0 ir T, atitinkamai, pirmoji ir paskutinioji amžiaus grupės, įtrauktos į skaičiavimus.

Rezultatai. Vyrams neparametriniu būdu įvertintas vidutinis kvadratinis nuokrypis (6) buvo lygus 0.0107404, moterims 0.0097395. Vertinant modelį (3) su neparametrine resurso funkcija (4) vyrams v.k.n. buvo lygus 0.0218859, moterims 0.0170842. Parametrinės resurso funkcijos (5) atveju v.k.n. buvo lygus 0.0265938 ir 0.00199838, atitinkamai.



2 pav. Laiko faktoriaus $x(i)$ reikšmės, kai $G(\Lambda)$ ir $\Lambda_k(t)$ parametrai įvertinti pagal Lietuvos ne jaunesnių kaip 40 metų amžiaus gyventojų 1980–1996 metų mirtingumo duomenis ($x(1996)$ prilygintas vienetui).

Literatūra

[1] D.J. Slimen, P.A. Lasenbruch, Survival distributions arising from two families and generated by transformations, *Commun. Statist.: Theory and Meth.*, **13**(10), 1179–1201 (1984).

[2] P. Hougaard, Fraitly models for survival data, *Lifetime Data Analysis*, **1**, 255–273 (1995).

[3] R. Mišeikytė, Lietuvos gyventojų kohortinio išgyvenamumo modeliavimas, *LMD mokslo darbai*, Vilnius, 366–371 (1998).

Application of generalized multiplicative reliability function to Lithuanian's survival data

R. Senkuvienė

Generalized multiplicative survival function was fitted to Lithuanian 1980–1996 years mortality data in 40+ age group. Quality of fit was evaluated using nonparametric variance estimation.