

Eva Kotlebová

BAYESOVSKÝ BODOVÝ ODHAD PARAMETRA POISSONOVHO ROZDELENIA

Úvod

Poissonovo rozdelenie je v štatistike pomerne často využívaným rozdelením, pretože sa ním dá modelovať široká škála premenných. Okrem toho má niektoré výhodné vlastnosti: má jednoduchú pravdepodobnostnú funkciu, možno ním aproximovať binomické rozdelenie a jeho jediný parameter je súčasne jeho strednou hodnotou aj disperziou. Jeho pozitívne zošikmenie sa so zväčšovaním hodnoty parametra znižuje.

V poisťovníctve sa toto rozdelenie štandardne využíva na modelovanie počtu poisťovních udalostí, čo je jednou z tých veličín, ktorých dôkladné poznanie je pre poisťovňu veľmi dôležité. Kľúčovou informáciou je predpokladaný priemerný počet poisťovních udalostí, ktorý sa odhaduje práve pomocou strednej hodnoty Poissonovho rozdelenia. V klasickej induktívnej štatistike je bodovým odhadom tohto parametra výberový priemer, ktorý z matematického hľadiska spĺňa všetky potrebné podmienky kladené na bodový odhad.

Pri krátkodobých poisťných zmluvách v neživotnom poistení považujeme za náhodný výber údaje z posledného obdobia. Vzhľadom na to, že poisťovňa často disponuje aj údajmi z predchádzajúcich období, prípadne inými relevantnými zdrojmi, je nanajvýš žiadúce zahrnúť ich do analýzy za účelom spresnenia odhadu. Tu sa ponúka priestor na využitie bayesovských metód.

Bayesovská štatistika využíva na riešenie induktívnych úloh dva druhy informácie: popri údajoch pochádzajúcich z výberového zisťovania berie do úvahy aj tzv. apriórnu informáciu, ktorá má podobu pravdepodobnostného rozdelenia odhadovaného parametra. Definovanie takéhoto rozdelenia je však často veľmi problematické; najmä stanovenie hodnôt jeho parametrov vyžaduje nielen poznanie daného problému, ale aj istú mieru abstrakcie. Ich deficit niekedy vedie k upusteniu od použitia bayesovských metód, takže núkajúce sa zdroje informácie ostanú nepovšimnuté, čo iste nemožno považovať za pozitívne.

Príspevok si kladie za cieľ ukázať spôsob, akým možno definovať apriórne rozdelenie na základe jednoduchej predstavy, pričom bodový odhad, ktorý je výsledkom navrhovaného algoritmu, má veľmi dobré matematické vlastnosti v porovnaní nielen s klasickým bodovým odhadom, ale aj so štandardne používaným bayesovským odhadom.

1 BODOVÝ ODHAD PARAMETROV V BAYESOVSKÉJ ŠTATISTIKE

Na rozdiel od klasickej štatistickej indukcie, v ktorej je odhadovaný parameter Θ neznámou konštantou, v bayesovskej štatistike je považovaný za náhodnú premennú. Jej rozdelenie je ovplyvnené jednak apriórnou predstavou, na základe ktorej sa formuje apriórne rozdelenie, ale aj výberovými údajmi, vplyvom ktorých sa apriórne rozdelenie

aktualizuje – vytvára sa tzv. aposteriórne rozdelenie, ktoré je podkladom pre indukčné závery. Vzťah medzi apriórnym, aposteriórnym a výberovým rozdelením je definovaný pomocou Bayesovej vety o podmienenej pravdepodobnosti, ktorej spojitá verzia má tvar [Pacáková, 2004]

$$f_{\Theta}(\theta | \mathbf{x}) = \frac{f(\mathbf{x} | \theta) \cdot f_{\Theta}(\theta)}{\int_{\Theta} f(\mathbf{x} | \theta) \cdot f_{\Theta}(\theta) d\theta}, \quad (1)$$

pričom

$f_{\Theta}(\theta)$ označuje apriórnu hustotu odhadovaného parametra Θ

$f_{\Theta}(\theta | \mathbf{x})$ označuje aposteriórnu hustotu parametra Θ

$f(\mathbf{x} | \theta)$ je funkcia vierohodnosti.

V menovateli zlomku je tzv. marginálna hustota $f(\mathbf{x})$ (integruje sa cez množinu možných hodnôt premennej Θ), ktorá nezávisí od konkrétnej hodnoty parametra Θ , preto sa zvykne zanedbávať a vzťah rovnosti sa nahradí vzťahom proporcionality

$$f_{\Theta}(\theta | \mathbf{x}) \propto f(\mathbf{x} | \theta) \cdot f_{\Theta}(\theta) \quad (2)$$

V prípade, že aposteriórne rozdelenie je rovnakého typu ako apriórne, nazývame ho konjugovaným rozdelením k rozdeleniu, z ktorého pochádza náhodný výber, pričom trojica rozdelení (apriórne, výberové a aposteriórne) tvorí tzv. konjugovaný systém (conjugate family). Pre niektoré konjugované systémy boli odvodené vzorce na výpočet hodnôt parametrov aposteriórneho rozdelenia, ktoré podstatne zjednodušujú potrebné výpočty aj celý algoritmus nájdenia odhadu. Je veľmi výhodné, ak apriórne (aj aposteriórne) rozdelenie patrí k dobre známym pravdepodobnostným rozdeleniam, ktoré sú bežnou súčasťou štatistických softvérov.

V poistnej praxi (ale aj v rôznych iných oblastiach) sa najčastejšie využívajú konjugované systémy binomické/beta (na odhad parametra π binomického rozdelenia), Poissonovo/gama (na odhad parametra λ Poissonovho rozdelenia) a normálne/normálne (na odhad strednej hodnoty μ normálneho rozdelenia).

Bayesovským bodovým odhadom parametra Θ je niektorá charakteristika aposteriórneho rozdelenia; najčastejšie používa jeho stredná hodnota $E(\theta | \mathbf{x})$.

Pre odhad parametra λ Poissonovho rozdelenia je výhodné za apriórne rozdelenie zvoliť rozdelenie Gama ($G(\alpha; \beta)$), pretože je dostatočne flexibilné a vhodnou voľbou číselných hodnôt parametrov umožňuje modelovanie širokého spektra apriórnych predstáv o parametri. Na základe vzťahu (2) bolo dokázané ([3], [7]), že pre parametre α', β' aposteriórneho gama-rozdelenia (premennú označujeme ako $\lambda | \mathbf{x}$, kde $(x_1, x_2, \dots, x_n) = \mathbf{x}$ je vektor zistených hodnôt) platí:

$$\alpha' = \alpha + \sum_{i=1}^n x_i \quad (3)$$

$$\beta' = \beta + n \quad (4)$$

Bayesovským bodovým odhadom parametra λ je stredná hodnota tohto rozdelenia:

$$E(\lambda / \mathbf{x}) = \frac{\alpha'}{\beta'} = \frac{\alpha + \sum_{i=1}^n x_i}{\beta + n} \quad (5)$$

2 STREDNÁ KVADRATICKÁ CHYBA AKO KRITÉRIUM KVALITY BODOVÉHO ODHADU

Všeobecne platí, že bodovým odhadom parametra Θ rozdelenia náhodnej premennej je výberová charakteristika U_n (zapisujeme $est\Theta = U_n$), ktorá spĺňa isté kritériá. Najdôležitejšie z nich sú:

- neskreslenosť (nevychýlenosť) – stredná hodnota charakteristiky U_n sa rovná odhadovanému parametru ($E(U_n) = \Theta$),
- konzistentnosť - so zväčšujúcim sa rozsahom výberového súboru by sa jej hodnota mala blížiť k odhadovanému parametru ($\lim_{n \rightarrow \infty} P(|U_n - \Theta| < \varepsilon) = 1$),
- výdatnosť – jej rozptyl je najmenší spomedzi rozptylov všetkých neskreslených odhadov.

Prvá z uvedených vlastností má pritom najvýznamnejšie postavenie, pretože podľa definície odhad nemôže byť výdatný, ak je čo len nepatrne skreslený. To znamená, že ak by sme dôsledne trvali na splnení všetkých troch podmienok, mohlo by sa stať, že by sme uprednostnili (síce) neskreslený odhad s vysokou variabilitou pred minimálne skresleným odhadom s nepatrnou variabilitou.

Rozumným prístupom je zohľadniť naraz obidve uvedené vlastnosti – najlepším odhadom by mala byť charakteristika s minimálnym rozptylom a súčasne s minimálnym skreslením. Z tohto hľadiska sa javí vhodnou mierou stredná kvadratická chyba odhadu

$$MSE(U_n) = E[(\Theta - U_n)^2] = D(U_n) + \Delta_n^2 \quad (6)$$

v ktorej $D(U_n)$ označuje rozptyl charakteristiky U_n a $\Delta_n = E(U_n) - \Theta$ je skreslenie odhadu.

Najlepším odhadom z tohto hľadiska je ten, ktorého stredná kvadratická chyba je najmenšia. Snažili sme sa nájsť bayesovský bodový odhad parametra λ , ktorý by bol v tomto zmysle lepší ako klasický bodový odhad, pričom sa nám podarilo nájsť aj algoritmus na optimálnu voľbu parametrov apriórneho rozdelenia na základe jednoduchej apriórnej predstavy.

Použili sme pri tom úvahy z [8] a [9], kde je problém vyriešený pre konjugovaný systém binomické/beta. Bolo ukázané, že na základe jednoduchej apriórnej predstavy o odhadovanom parametri π (príslušnosť ku konkrétnemu intervalu) možno stanoviť parametre apriórneho rozdelenia tak, že bayesovský bodový odhad má nižšiu kvadratickú chybu ako klasický odhad práve na stanovenom intervale.

Cieľom tohto príspevku je nájsť podobný algoritmus pre určenie bodového odhadu parametra λ Poissonovho rozdelenia.

3 POROVNANIE STREDNÝCH KVADRATICKÝCH CHÝB BAYESOVSKÉHO A KLASICKÉHO BODOVÉHO ODHADU PARAMETRA POISSONOVHO ROZDELENIA

Ako je všeobecne známe, klasickým bodovým odhadom parametra Poissonovho rozdelenia je výberový priemer: $est\lambda = \bar{x}$. Je to odhad neskreslený, preto je jeho stredná kvadratická chyba rovná rozptylu

$$MSE(\bar{X}) = D(\bar{X}) + 0^2 = \frac{\lambda}{n}. \quad (7)$$

Bayesovským bodovým odhadom parametra λ (označujeme $\hat{\lambda}_B$) je stredná hodnota aposteriórneho rozdelenia $G(\alpha + \sum_{i=1}^n x_i; \beta + n)$

$$\hat{\lambda}_B = E(\lambda / \mathbf{x}) = \frac{\alpha + n\bar{x}}{\beta + n}. \quad (8)$$

Aby sme mohli vyjadriť strednú kvadratickú chybu tohto odhadu, určíme jeho (prípadné) skreslenie $(E(\hat{\lambda}_B) - \lambda)$ a rozptyl $D(\hat{\lambda}_B)$. Stredná hodnota bayesovského odhadu má tvar

$$E(\hat{\lambda}_B) = \frac{1}{\beta + n} [E(\alpha) + nE(\bar{X})] = \frac{\alpha + n\lambda}{\beta + n} \neq \lambda, \quad (9)$$

z ktorého je zrejmé, že ide o skreslený odhad. Veľkosť skreslenia je rozdiel

$$\frac{\alpha + n\lambda}{\beta + n} - \lambda = \frac{\alpha - \beta\lambda}{\beta + n}. \quad (10)$$

Rozptyl bayesovského odhadu môžeme vyjadriť nasledovne:

$$D(\hat{\lambda}_B) = \frac{1}{(\beta + n)^2} [D(\alpha) + n^2 D(\bar{X})] = \frac{1}{(\beta + n)^2} [0 + n\lambda] = \frac{n\lambda}{(\beta + n)^2} . \quad (11)$$

Po dosadení (10) a (11) do (6) dostaneme strednú kvadratickú chybu bayesovského odhadu v tvare

$$MSE(\hat{\lambda}_B) = \frac{1}{(\beta + n)^2} [(\alpha - \beta\lambda)^2 + n\lambda] = \frac{1}{(\beta + n)^2} [\beta^2\lambda^2 + \lambda(n - 2\alpha\beta) + \alpha^2] . \quad (12)$$

Obidve vyjadrené stredné kvadratické chyby (klasického (7) a bayesovského bodového odhadu(8)) obsahujú vo svojom vyjadrení rôzne premenné. My sme k nim pristupovali ako k funkcii premennej λ . Z tohoto hľadiska je stredná kvadratická chyba klasického odhadu (7) lineárnou funkciou a stredná kvadratická chyba bayesovského odhadu (12) konvexnou kvadratickou funkciou. Naším cieľom bolo nájsť taký interval, na ktorom je stredná kvadratická chyba bayesovského bodového odhadu menšia ako stredná kvadratická chyba klasického bodového odhadu, t. j. na ktorom platí

$$MSE(\hat{\lambda}_B) < MSE(\bar{X}) . \quad (13)$$

Ak by sa nám podarilo nájsť nejakú hodnotu λ , pre ktorú by táto podmienka platila, musel by existovať aj interval hodnôt s touto vlastnosťou. (Časť paraboly by ležala pod priamkou.)

Ukázalo sa, že ak by stredná hodnota apriórneho rozdelenia bola rovná hodnote λ ($\lambda = \frac{\alpha}{\beta}$), tak by požadovaná nerovnosť platila:

$$\begin{aligned} MSE(\hat{\lambda}_B) &= \frac{1}{(\beta + n)^2} \left[\left(\alpha - \beta \cdot \frac{\alpha}{\beta} \right)^2 + n \frac{\alpha}{\beta} \right] = \frac{n\alpha}{\beta(\beta + n)^2} = \\ &= \frac{\alpha}{\beta n} \cdot \frac{n^2}{(\beta + n)^2} < \frac{\alpha}{\beta n} (= MSE(\bar{X}) \text{ pre } \lambda = \frac{\alpha}{\beta}) \end{aligned}$$

Je teda zrejmé, že existuje interval $(\lambda_1; \lambda_2)$ obsahujúci bod $\frac{\alpha}{\beta}$, na ktorom nerovnosť (13) platí. Výhodou by bolo, ak by to bol interval zodpovedajúci apriórnej predstave o odhadovanom parametre λ .

Hranicami intervalu sú priesečníky grafov funkcií stredných kvadratických chýb porovnávaných druhov odhadov. Pre hodnoty $\lambda_1; \lambda_2$ teda platia vzťahy (14) a (15).

$$\frac{1}{(\beta + n)^2} [(\alpha - \beta\lambda_1)^2 + n\lambda_1] = \frac{\lambda_1}{n}, \quad (14)$$

$$\frac{1}{(\beta + n)^2} [(\alpha - \beta\lambda_2)^2 + n\lambda_2] = \frac{\lambda_2}{n} \quad (15)$$

Využijeme ich na určenie parametrov takého apriórneho rozdelenia, ktoré vedie k bayesovskému bodovému odhadu s požadovanými vlastnosťami. Postup riešenia nie je náročný, ale je dosť zdĺhavý, preto uvidíme len niektoré kroky. Najskôr po úprave rovníc vyjadríme premennú α

$$\alpha = \frac{1}{2n} [\beta n(\lambda_1 + \lambda_2) - \beta - 2n] \quad (16)$$

a dosadíme ju do rovnice (14). Dostaneme kvadratickú rovnicu s neznámou β

$$\begin{aligned} n \left\{ \frac{1}{2n} [\beta n(\lambda_1 + \lambda_2) - \beta - 2n] \right\}^2 - 2n \frac{1}{2n} [\beta n(\lambda_1 + \lambda_2) - \beta - 2n] \beta \lambda_1 + n \beta^2 \lambda_1^2 = \\ = \beta^2 \lambda_1 + 2n \beta \lambda_1, \end{aligned} \quad (17)$$

ktorú upravíme na tvar

$$\beta^2 [n^2 (\lambda_1 - \lambda_2)^2 - 2n(\lambda_1 + \lambda_2) + 1] + \beta [-4n^2 (\lambda_1 + \lambda_2) + 4n] + 4n^2 = 0 \quad (18)$$

Diskriminant D je kladný

$$D = 64n^4 \lambda_1 \lambda_2 > 0, \quad (19)$$

preto dostaneme dve riešenia:

$$\beta_{12} = \frac{4n[n(\lambda_1 + \lambda_2) - 1] \pm 8n^2 \sqrt{\lambda_1 \lambda_2}}{2[n^2 (\lambda_2 - \lambda_1)^2 - 2n(\lambda_1 + \lambda_2) + 1]} = \frac{2n[n(\lambda_1 + \lambda_2) - 1] \pm 4n^2 \sqrt{\lambda_1 \lambda_2}}{[n^2 (\lambda_2 - \lambda_1)^2 - 2n(\lambda_1 + \lambda_2) + 1]} \quad (20)$$

(ku každému z nich dopočítame α podľa (16)).

V prípade, že vo vyjadrení β podľa (20) zvolíme možnosť „-“, premenná α bude mať záporné znamienko. Naše (v praxi využiteľné) riešenie teda vyzerá nasledovne:

$$\beta = \frac{2n[n(\lambda_1 + \lambda_2) - 1] + 4n^2 \sqrt{\lambda_1 \lambda_2}}{[n^2(\lambda_2 - \lambda_1)^2 - 2n(\lambda_1 + \lambda_2) + 1]}, \quad (21)$$

$$\alpha = \frac{1}{2n} [\beta n(\lambda_1 + \lambda_2) - \beta - 2n]. \quad (22)$$

Prínos uvedených vzťahov spočíva v možnosti veľmi jednoduchého stanovenia parametrov apriórneho rozdelenia, pretože v praxi sa práve táto etapa javí ako dosť veľký problém. Ak by sme vedeli spoľahlivo určiť strednú hodnotu a rozptyl apriórneho rozdelenia, jeho parametre by sa určili riešením sústavy

$$E(\lambda) = \frac{\alpha}{\beta}, \quad (23)$$

$$D(\lambda) = \frac{\alpha}{\beta^2} \quad (24)$$

Kým stredná hodnota sa zvyčajne dá určiť pomerne presne, predstavy o variabilite sú o niečo hmlistejšie. Preto aj výsledok riešenia sústavy rovníc (23) a (24) treba brať s určitou rezervou, pričom tieto pochybnosti sa prenášajú aj do zodpovedajúceho aposteriórneho rozdelenia a z neho vyplývajúceho odhadu.

Preto sa javí oveľa jednoduchšie vyjadriť apriórnu predstavu v tvare nejakého intervalu $(\lambda_1; \lambda_2)$, v ktorom by sa odhadovaný parameter λ mal nachádzať. Ak by sme takú predstavu mali a popri tom by sme poznali aj rozsah výberového súboru n , mohli by sme z rovníc (21) a (22) vyjadriť parametre apriórneho rozdelenia a následne stanoviť bayesovský odhad.

Treba však zdôrazniť, že rovnice (14) a (15), ktorých riešením sú vzťahy (21) a (22), boli vytvorené tak, aby bolo zaručené, že na intervale $(\lambda_1; \lambda_2)$ dostaneme taký bayesovský bodový odhad, ktorého stredná kvadratická chyba je v porovnaní s klasickým bodovým odhadom menšia (teda je to kvalitnejší odhad).

Prínos uvedených vzťahov je teda dvojaký:

- zaručujú jednoduché určenie apriórneho rozdelenia,
- výsledný bayesovský bodový odhad parametra λ má lepšie vlastnosti (menšiu kvadratickú chybu) ako klasický bodový odhad.

Na ilustráciu odvodených vzťahov a s nimi súvisiacich vlastností sme zvolili hodnoty vstupných premenných: $\lambda_1 = 8$, $\lambda_2 = 13$, $n = 25$.

Najskôr sme pomocou vzťahov (21) a (22) vypočítali hodnoty parametrov apriórneho rozdelenia: $\beta = 3,546590$ $\alpha = 36,168264$. Vypočítané hodnoty sme využili na vyjadrenie stredných kvadratických chýb porovnávaných odhadov.

Pre klasický bodový odhad platí

$$MSE(\lambda) = \frac{\lambda}{n} = \frac{\lambda}{25}. \quad (25)$$

Pre vyjadrenie strednej kvadratickej chyby bayesovského bodového odhadu sme dosadili do (12):

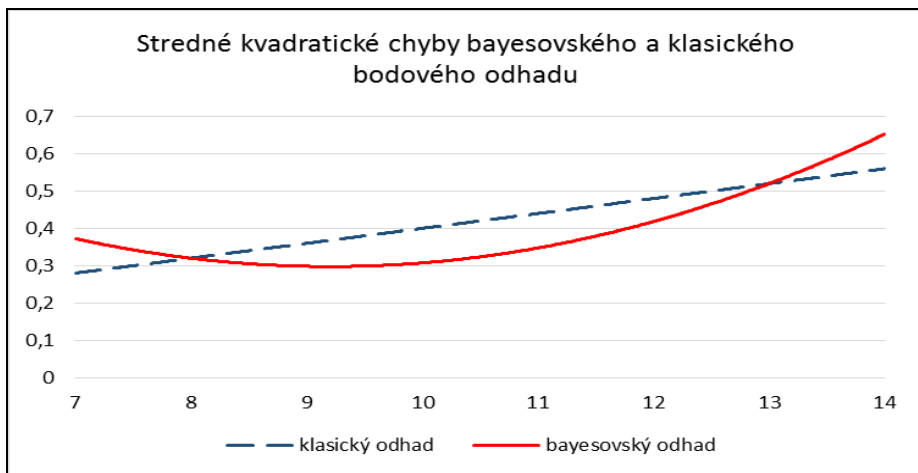
$$MSE(\hat{\lambda}_B) = \frac{1}{(3,546590 + 25)^2} \cdot [3,546590^2 \lambda^2 + \lambda \cdot (25 - 2 \cdot 36,168264 \cdot 3,546590) + 36,168264^2].$$

Po zjednodušení sme dostali kvadratickú funkciu

$$MSE(\hat{\lambda}_B) = 0,015435\lambda^2 - 0,284140\lambda + 1,605265$$

Obidve funkcie ((25) a (26)) sme znázornili do jedného grafu – zamerali sme sa na interval $\langle 7;14 \rangle$, ktorý pokrýva apriórne predpokladaný interval $\langle 8;13 \rangle$. Grafy sú na obrázku 1.

Obr.1: Porovnanie stredných kvadratických chýb klasického a bayesovského bodového odhadu parametra Poissonovho rozdelenia



Ako vidíme, grafy funkcií sa pretínajú v bodoch, ktorých x-ové súradnice zodpovedajú apriórne predpokladaným hraniciam, pričom vnútri intervalu $\langle 8;13 \rangle$ nadobúda stredná kvadratická chyba bayesovského bodového odhadu menšie hodnoty ako stredná kvadratická chyba klasického bodového odhadu. Z matematického hľadiska tak ide o kvalitnejší bodový odhad. Mimo intervalu je situácia opačná – stredná kvadratická chyba je väčšia, pričom rozdiel oproti strednej kvadratickej chybe klasického odhadu sa so zväčšujúcou vzdialenosťou od hraníc intervalu významne zvyšuje. Preto je nutné, aby

hranice intervalu $(\lambda_1; \lambda_2)$ boli stanovené čo najpresnejšie, v prípade pochybností treba zvoliť radšej širší interval.

Hodnota bayesovského bodového odhadu (a jeho poloha v súvislosti s hranicami intervalu) závisí aj od výberových údajov, konkrétne od charakteristiky $\sum_{i=1}^n x_i$.

Teoreticky môže dôjsť aj k situácii, že bayesovský odhad leží mimo apriórne predpokladaného intervalu $(\lambda_1; \lambda_2)$, čo jeho kvalitu významne znižuje. Ide o prípad, keď výberový priemer leží mimo intervalu $(\lambda_1; \lambda_2)$, čo však svedčí o dosť pochybnnej a od praxe vzdialenej apriórnej predstave.

Prezentované vzťahy a súvislosti medzi nimi nám umožnili navrhnúť algoritmus na určenie bodového odhadu parametra λ Poissonovho rozdelenia na základe jednoduchšej apriórnej predstavy:

1. stanovenie hraníc intervalu $(\lambda_1; \lambda_2)$, zodpovedajúcich apriórnej predstave o parametri λ
2. vyčíslenie hodnôt parametrov α, β apriórneho rozdelenia gama podľa vzťahov

$$\beta = \frac{2n[n(\lambda_1 + \lambda_2) - 1] + 4n^2\sqrt{\lambda_1\lambda_2}}{[n^2(\lambda_2 - \lambda_1)^2 - 2n(\lambda_1 + \lambda_2) + 1]}$$

$$\alpha = \frac{1}{2n}[\beta n(\lambda_1 + \lambda_2) - \beta - 2n]$$

3. určenie hodnoty bayesovského bodového odhadu parametra λ pomocou vzťahu

$$\hat{\lambda}_B = E(\lambda / \mathbf{x}) = \frac{\alpha + n\bar{x}}{\beta + n}$$

Uvedený algoritmus umožňuje stanovenie kvalitného bayesovského bodového odhadu parametra λ Poissonovho rozdelenia, pričom nie je nutné, aby ten, kto ho používa, ovládal teoretické princípy bayesovskej štatistiky.

4 MOŽNOSTI APLIKÁCIE ALGORITMU

Ako už bolo uvedené, algoritmus bol vyvinutý na základe potreby čo najpresnejšie odhadnúť priemerný počet poistných udalostí v nasledujúcom období v krátkodobých poistných zmluvách. Doterajší spôsob využitia bayesovských metód bol postavený na definícii apriórneho gama rozdelenia s konkrétnymi hodnotami parametrov, čo predpokladalo určitú mieru abstrakcie. Navrhovaný spôsob určenia apriórneho rozdelenia je skutočne jednoduchý, odhadnúť hranice intervalu, v ktorom sa odhadovaný parameter

má nachádzať, dokáže aj laik, ktorý z dostupných údajov (podkladov pre apriórne rozdelenie) zvolí najmenšiu hodnotu ako dolnú hranicu λ_1 a najväčšiu hodnotu ako λ_2 .

Okrem poisťovníctva je možné opísaný postup aplikovať všade tam, kde potrebujeme odhadnúť priemernú hodnotu premennej, ktorá sa dá modelovať Poissonovým rozdelením. Vhodnou oblasťou je napríklad demografia podnikov, kde základné ukazovatele - vznik a zánik podnikov, prípadne zlučovanie alebo delenie je možné modelovať týmto rozdelením a odhadovať priemerný počet udalostí v nasledujúcich rokoch. Odhady, ktoré takýmto spôsobom vzniknú, môžu byť cenným podkladom pre inštitúcie tvoriace hospodársku politiku.

Záver

Vo všeobecnosti sa bayesovský prístup pri riešení indukčných úloh nevyužíva tak často, ako by bolo žiadúce, napriek tomu, že umožňuje pri analýze využiť viaceré zdroje informácie. Príčinou je jednak náročnejší matematický aparát, ktorý tento prístup vyžaduje, ale aj problematické stanovenie apriórneho rozdelenia. Apriórne informácie pochádzajúce z rôznych zdrojov sa niekedy dosť významne líšia, pričom niekedy majú subjektívny charakter, čo môže spochybniť závery, ktoré sú na základe takéhoto vstupu vyvedené.

Ukázali sme, že pri podrobnej analýze vzťahov medzi premennými zastúpenými v konjugovanom systéme Poissonovo/gama je možné dať ich do súvislostí, ktoré prinesú pozitívny a v praxi využiteľný efekt, pričom odhad, ktorý na základe prezentovaného algoritmu vznikne, má dobré matematické vlastnosti. Využitie strednej kvadratickej chyby ako kritéria kvality bodového odhadu ponúka komplexnejší pohľad na problematiku a z neho vyplývajúce možnosti aplikácie.

Kľúčová slová

Bayesovská štatistika, bodový odhad, Poissonovo rozdelenie, apriórne rozdelenie, aposteriórne rozdelenie, stredná kvadratická chyba, počet poisťných udalostí, konjugované rozdelenie

Klasifikácia JEL

C11

LITERATÚRA

- [1] BAKYTOVÁ, H. – HÁTLE, J. – NOVÁK, I. – UGRON, M. 1986. *Statistická indukce pro ekonomy*. Praha : SNTL, ALFA, 1986.
- [2] BERNARDO, JOSÉ M. – SMITH, ADRIAN F. M. 1995. *Bayesian theory*. Chichester : John Wiley & Sons Ltd., 1995.
- [3] BOLSTAD, W.M. 2004. *Introduction to Bayesian statistics*. New Jersey, USA : John Wiley&sons, Inc. 2004.

- [4] FREUND, J. E. 1992. *Mathematical statistics*. New Jersey : Prentice-Hall, International, 1992.
- [5] GARTHWAITE, P. H. – JOLLIFFE, I. – JONES, B. 1995. *Statistical Inference*. Prentice-Hall International, Inc., 1995.
- [6] HORÁKOVÁ, G. – HUŤKA, V.: *Teória pravdepodobnosti I*. Bratislava : Vyd. EKONÓM, 2002.
- [7] KOTLEBOVÁ, E. 2009. *Bayesovská štatistická indukcia v ekonomických aplikáciách*. Bratislava : Vyd. EKONÓM, 2009.
- [8] KOTLEBOVÁ, E. – LÁSKA, I. 2014. *Bayesovský prístup k bodovému odhadu pravdepodobnosti poistnej udalosti*. Bratislava : Ekonomika a informatika, 1/2014, s. 101 – 112, ISSN 1336-3514.
- [9] KOTLEBOVÁ, E.- LÁSKA I. 2014. *Využitie bayesovského prístupu pri odhade podielu a možnosti jeho aplikácie v ekonomickej praxi*. Bratislava : Slovenská štatistika a demografia 2/2014, s. 3 – 17.
- [10] LEE, PETER M. 1989. *Bayesian statistics*. New York : Oxford University Press, 1989.
- [11] PACÁKOVÁ, V. 2004. *Aplikovaná poistná štatistika*. Bratislava : IURA EDITION, 2004.
- [12] PACÁKOVÁ, V. a kol. 2012. *Štatistická indukcia pre ekonómov*. Bratislava : IURA EDITION, 2004.
- [13] PFAFFENBERGER, R. C. – PATTERSON, J. H. 1987. *Statistical Methods (For business and Economics)*. Illinois : IRWIN, 1987.
- [14] ŠOLTĚS, E. 2009. *Modely kredibility na výpočet poistného*, Bratislava: Vyd. EKONÓM, 2009.
- [15] TEREK, M. 2003. *Úvod do analýzy rozhodovania a bayesovskej indukcie*. Bratislava : Vyd. EKONÓM, 2003.

RESUMÉ

Príspevok sa zaoberá bayesovským bodovým odhadom parametra Poissonovho rozdelenia, na ktoré sa v bayesovskej štatistike bežne využíva konjugovaný systém Poissonovo/gama. Prezentovaný prístup je tiež postavený na tomto konjugovanom systéme, ale odlišuje sa v tom, že ako kritérium kvality bodového odhadu sa používa stredná kvadratická chyba.

Podrobná analýza súvislostí medzi zastúpenými premennými viedla jednoduchému algoritmu na stanovenie bayesovského bodového odhadu, pričom ako prínos sa javí fakt, že algoritmus umožňuje veľmi jednoduchý spôsob stanovenia parametrov apriórneho gama rozdelenia. Výsledkom je taký bayesovský bodový odhad, ktorého stredná kvadratická chyba je menšia v porovnaní s klasickým odhadom na vopred stanovenom intervale, ktorý zodpovedá apriórnej predstave.

SUMMARY

The contribution deals with the bayesian point estimation of the Poissons' parameter, for which the conjugate family Poisson/gamma is usually used in bayesian statistics. Presented approach is based on this conjugate family, too, but differs in the criterion of point estimations' quality - in the contribution the mean square error is used.

Detailed analysis of relations among the variables occurred in model allowed to develop the algorithm for determination the bayesian point estimation. As a great benefit the possibility of simple definition of prior gamma distributions' parameters appears. The consequence is such bayesian point estimation, the mean square error of which is smaller than that of the classical point estimation within the pre-determined interval, which corresponds with the prior conception of estimated parameter.

Kontakt

RNDr. Eva Kotlebová, PhD., Katedra štatistiky, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave, Dolnozemska cesta 1, 852 35 Bratislava, tel.: +421 2/672 95 718, e-mail: eva.kotlebova@gmail.com