ŽILINSKÁ UNIVERZITA V ŽILINE
FAKULTA ŠPECIÁLNEHO INŽINIERSTVA

KRÍZOVÝ MANAŽMENT - 1/2011

MODELOVANIE FREKVENCIE DOPRAVNÝCH NEHÔD V SKUPINE VODIČOV

Aleš KOZUBÍK [1]

SUMMARY

In the present article, we analyse several models of the frequency of road accidents. It is shown that the commonly used Poisson distribution is appropriate only for an individual loss process, but fails when applied to large portfolio of drivers. Therefore three additional claim frequency distributions are studied. It is shown that Poisson-Inverse Gaussian distribution and generalized Poisson fit well the empirical values.

ÚVOD

Jedným z najdôležitejších prvkov pri modelovaní dopravných nehôd sú, tak ako nakoniec pri všetkých modeloch krízových situácií, modely frekvencie ich výskytu. Vo všetkých prípadoch ide o pravdepodobnostné modely popisujúce rozdelenie počtu takýchto udalostí. Z poistnej praxe vyspelých krajín sú známe rôzne prístupy ku klasifikácii vodičov. Typickými sledovanými premennými sú napr. vek, pohlavie, profesia, bydlisko a typ používaného vozidla. V niektorých prípadoch sú to aj špecifickejšie premenné, ktoré by mali charakterizovať správanie vodiča, ako sú napr. rodinný stav, fajčenie či farba vozidla. Dlhodobé skúsenosti (napr. [6] alebo [7]) však ukazujú, že najlepším prediktorom budúcich nehôd daného vodiča je historický záznam jeho predchádzajúcich nehôd.

V tomto príspevku bude predstavených niekoľko pravdepodobnostných modelov, ktoré slúžia na modelovanie počtu dopravných nehôd.

1. POISSONOV MODEL

Poissonov model je založený na troch základných predpokladoch. Prvým je požiadavka, aby počet nehôd v určitom časovom intervale bol závislý len od dĺžky tohto časového intervalu (teda najmä nezávislý od jeho začiatku a konca). Druhou požiadavkou je, že v danom časovom intervale nemôže súčasne nastať viac ako jedna udalosť. Tretím predpokladom je nezávislosť počtu nehôd v disjunktných časových intervaloch. Je všeobecne známe, že pri rešpektovaní týchto troch predpokladov je počet dopravných nehôd počas roka charakterizovaný Poissonovým rozdelením pravdepodobnosti (pozri napr. [1] alebo [2]). Pre pravdepodobnosť p_k, že počet nehôd bude $k = 0,1,...$ platí

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!},\ k = 0,1,...,$$

kde λ je parameter Poissonovho rozdelenia. Jeho hodnota udáva strednú hodnotu a rozptyl tohto rozdelenia a možno ju teda interpretovať ako priemerný počet nehôd jedného vodiča.

Tabuľka 1
Empirické a teoretické početnosti poistných udalostí pri použití Poissonovho rozdelenia s parametrom odhadnutým momentovou metódou.

k	n_k	Poissonovo rozdelenie – Teoretické hodnoty (Parameter odhadnutý momentovou metódou)
0	3739	3699.8579
1	225	288.5889
2	24	11.2549
3	10	0.2926
4	1	0.0057
5	1	0
≥6	0	0

Zdroj: Vlastný zber dát s poisťovní v roku 2004 a vlastné výpočty.

[1] Katedra matematických metód, Fakulta riadenia a informatiky Žilinskej Univerzity v Žiline, Veľký Diel, 010 26 Žilina
e-mail: alesko@frcatel.fri.uniza.sk

Poissonovo rozdelenie aplikujeme na množinu reálnych dát reprezentujúcich 4000 poistných zmlúv na poistenie motorových vozidiel. Dáta rozdelíme do tried, podľa počtu nehôd u jednotlivých zmlúv. Parameter rozdelenia odhadneme momentovou metódou ako aritmetický priemer, teda $\lambda = 0{,}078$. Tak dostaneme výsledky podľa tabuľky č.1.

Už na prvý pohľad je zrejmé, že výsledky, získané pri použití Poissonovho rozdelenia sú nevyhovujúce. Navyše, k najväčším odchýlkam od skutočných početností dochádza práve pri hodnotách, ktoré sú z hľadiska frekvencie nehodovosti najvýznamnejšie.

Nevhodnosť aproximácie Poissonovým rozdelením možno overiť aj štatistickými testami. Bežne požívaným nástrojom na test zhody empirických údajov s predpokladaným rozdelením je χ^2-testom dobrej zhody. Pre testovaciu štatistiku

$$\chi^2 = \sum_{k=0}^{m} \frac{\left(n_k - np_k\right)^2}{np_k}$$

dostávame hodnotu *150,2987*, čo je hodnota výrazne prevyšujúca kritické hodnoty rozdelenia χ^2 s piatimi stupňami voľnosti, ktoré sú pre spoľahlivosť 95% $\chi^2_{5;0,95} = 11{,}07$ a pre spoľahlivosť 99,5% $\chi^2_{5;0,995} = 16{,}75$. Tieto výsledky potvrdzujú jednoznačné zamietnutie hypotézy o Poissonovom rozdelení pravdepodobnosti počtu dopravných nehôd.

Vzhľadom na nízky počet tried v uvedenom súbore a nízky počet pozorovaní vo viacerých triedach, nemusí byť χ^2-test dostatočne preukazný. (Obvykle sa požaduje viac ako 5 tried a aspoň 5 pozorovaní v každej triede, niekedy dokonca až 10 pozorovaní v každej triede.) Preto tento výsledok overíme ešte tzv. G-testom (uvádzaným napr. v [8]), ktorý vyžíva presnejšiu aproximáciu multinoimckého rozdelenia rozdelením χ^2 pri nižšom počte tried. Jeho testovacia štatistika má tvar

$$G = 2\sum_{k=0}^{m} n_k \ln\left(\frac{n_k}{np_k}\right).$$

Táto náhodná premenná má potom opäť rozdelenie χ^2 s rovnakým počtom stupňov voľnosti ako pri χ^2-teste. Pre Poissonovo rozdelenie pre hodnotu G štatistiky dostávame 17,66 čo opäť prevyšuje obe kritické hodnoty.

2. NEGATÍVNE BINOMICKÝ MODEL

Zamietnutie Poissonovho modelu znamená, že správanie vodičov je heterogénne. Pre modelovanie frekvencie nehodovosti v rámci veľkej skupiny vodičov je preto potrebné hľadať rozdelenia, ktoré by odrážali túto heterogenitu v správaní. Jednou z ciest je predpokladať, že počet nehôd u jednotlivého vodiča sa riadi Poissonovým rozdelením, avšak jeho parameter λ sa u každého jednotlivého vodiča mení. Pri tomto prístupe považujeme parameter λ za konkrétnu realizáciu náhodnej premennej Λ. Ak budeme predpokladať, že náhodná premenná Λ je spojitá a označíme jej funkciu hustoty ako $u(\lambda)$, tak pre výsledné rozdelenie počtu nehôd (napr. podľa [2]) dostávame

$$p_k = \int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!} u(\lambda)\,\mathrm{d}\lambda, \quad k = 0,1,....$$

Takýto typ rozdelení nazývame zmiešanými Poissonovými rozdeleniami a rozdelenie náhodnej premennej Λ nazývame tzv. zmiešavacím rozdelením. (Bližšie o zmesiach rozdelení napr. v [2].)

Ak za zmiešavacie rozdelenie zvolíme gama rozdelenie s parametrami a a τ, ktoré má hustotu tvaru

$$u(\lambda) = \frac{\tau^a e^{-\tau\lambda}\lambda^{a-1}}{\Gamma(a)}, \quad a, \tau > 0,$$

tak pre pravdepodobnosti počtu dopravných nehôd dostávame

$$p_k = \int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!} \frac{\tau^a e^{-\tau\lambda}\lambda^{a-1}}{\Gamma(a)}\,\mathrm{d}\lambda,$$

čo je možné upraviť do tvaru

$$p_k = \binom{k+a-1}{k} p^a q^k,$$

pričom kladieme

$$p = \frac{\tau}{1+\tau},\ q = 1 - p = \frac{1}{1+\tau}.$$

Zovšeobecnené binomické koeficienty definujeme pomocou Γ-funkcie ako

$$\binom{k+a-1}{k} = \frac{\Gamma(k+a)}{\Gamma(k+1)\Gamma(a)}.$$

Ako výsledok obdržíme negatívne binomické rozdelenie so strednou hodnotou $\frac{a}{\tau}$ a rozptylom $\sigma^2 = \frac{a}{\tau}\left(1+\frac{1}{\tau}\right)$.

To poukazuje na významnú vlastnosť negatívne binomického rozdelenia, ktorou je skutočnosť, že jeho rozptyl prevyšuje strednú hodnotu. Viac o tomto rozdelení napr. v [1] a [2].

Pre odhad parametrov negatívne binomického rozdelenia z momentovej metódy dostávame

$$\hat{\tau} = \frac{\bar{x}}{s^2 - \bar{x}}, \quad \hat{a} = \frac{\bar{x}^2}{s^2 - \bar{x}}.$$

Pri aplikácii na údaje z tabuľky číslo 1 dostávame odhady $\tau=2,697469$ a $a=0,210403$. Celkové výsledky sú potom zhrnuté v tabuľke číslo 2. Pre hodnotu testovacej štatistiky χ^2 - testu máme hodnotu 6,94, ktorá už neumožňuje predpokladané rozdelenie zamietnuť.

Podobne ako v predchádzajúcom odstavci, overíme ešte hodnotu G-štatistiky, pre ktorú v tomto prípade máme hodnotu 6,88.

3. POISSON - INVERZNÝ GAUSSOV MODEL

Podobne ako v predchádzajúcom odstavci predpokladáme, že nehody v heterogénnej množine vodičov je možné popísať zmiešaným Poissonvým rozdelením. O náhodnej premennej Λ však budeme predpokladať, že sa riadi inverzným Gaussovým rozdelením s funkciou hustoty (pozri [7])

$$u(\lambda) = \frac{g}{\sqrt{2\pi h \lambda^{\frac{3}{2}}}} e^{-\frac{1}{2h\lambda}(\lambda - g)^2}, \quad g,h > 0.$$

Výsledkom je potom zmiešané Poissonovo rozdelenie, ktoré sa nazýva Poissonovo – Inverzné Gaussovo rozdelenie. Jeho stredná hodnota nadobúda hodnotu g a jeho rozptyl je $\sigma^2 = g(1+h)$. Jednotlivé pravdepodobnosti p_k je možné vyjadriť tiež rekurentne v tvare

$$p_0 = e^{\frac{g}{h}[1-(1+2h)^{\frac{1}{2}}]},$$

$$p_1 = g p_0 (1+2h)^{-\frac{1}{2}},$$

Tabuľka 2

Empirické a teoretické početnosti poistných udalostí pri použití negatívneho binomického rozdelenia s parametrami odhadnutými momentovou metódou.

k	n_k	Teoretické hodnoty (Parameter odhadnutý momentovou metódou)
0	3739	3743,21
1	225	213,01
2	24	34,87
3	10	6,95
4	1	1,51
5	1	0,34
≥6	0	0,11

Zdroj: Vlastný zber dát s poisťovní v roku 2004 a vlastné výpočty.

Tabuľka 3

Empirické a teoretické početnosti poistných udalostí pri použití Poissonovho – inverzného Gaussovho rozdelenia s parametrami odhadnutými momentovou metódou.

k	n_k	Teoretické hodnoty (parameter odhadnutý momentovou metódou)
0	3739	3739,84
1	225	221,05
2	24	30,06
3	10	6,53
4	1	1,75
5	1	0,52
≥6	0	0,25

Zdroj: Vlastný zber dát s poisťovní v roku 2004 a vlastné výpočty.

$$(1+2h)k(k-1)p_k = h(k-1)(2k-3)p_{k-1} + \\ + g^2 p_{k-2}, k = 2,3,...$$

Pre odhady parametrov momentovou metódou potom platí

$$\hat{g} = \bar{x}, \hat{h} = \left(\frac{s^2}{\bar{x}}\right) - 1.$$

Pri aplikovaní na našu množinu 4000 údajov dostávame odhady $g=0{,}078$ a $h=0{,}370718$. Celkové výsledky pri použití Poissonovho-inverzného Gaussovho rozdelenie sú zhrnuté v tabuľke číslo 3. Pre hodnotu testovacej štatistiky χ^2 -testu máme hodnotu *3,897*, ktorá nielen že neumožňuje predpokladané rozdelenie zamietnuť, ale ako uvidíme je najlepšia zo všetkých testovaných rozdelení. Pre hodnotu G-testu dostávame taktiež veľmi priaznivú hodnotu 4,1973. Táto hodnota je výrazne nižšia, než tomu bolo pri negatívnom binomickom rozdelení, čo znamená, že toto rozdelenie nie je možné vylúčiť ani pri nižších hladinách spoľahlivosti..

4. ZOVŠEOBECNENÝ POISSONOV MODEL

Hovoríme, že náhodná premenná má zovšeobecnené Poissonovo rozdelenie (napr. [3] alebo [4]), ak jej pravdepodobnostná funkcia je daná vzťahom

$$p_k = \begin{cases} \lambda(\lambda + k\Theta)^{k-1} \dfrac{\exp(-\lambda - k\Theta)}{k!}, & k = 0,1,2,\dots \\ 0, \quad k > m & \text{ak} \quad \Theta < 0 \end{cases}$$

a 0 inak, kde $\lambda>0$, $\max(-1,-\lambda/m)\leq\Theta<1$ a $m\geq4$ je najväčšie prirodzené číslo, pre ktoré $\lambda+\Theta m>0$ ak Θ je záporné. V prípade, že $\Theta=0$ sa zovšeobecnené Poissonovo rozdelenie redukuje na bežné Poissonovo rozdelenie.

Pre určenie momentov zovšeobecneného Poissonovho rozdelenia využijeme momentovú vytvárajúcu funkciu, ktorú je možné vyjadriť v tvare

$$M(t) = \exp\left\{-\frac{\lambda}{\Theta}\left[W(-\Theta\exp(-\Theta+t)) + \Theta\right]\right\},$$

kde W je tzv. Lambertova funkcia, ktorú je možné definovať vzťahom

$$W(x)\exp(W(x)) = x.$$

Derivovaním momentovej vytvárajúcej funkcie môžeme určiť momenty zovšeobecneného Poissonovho rozdelenia. Tak určíme prvé štyri momenty

$$v_1 = \lambda M,$$
$$v_2 = \lambda M^3,$$
$$v_3 = \lambda(3M - 2)M^4,$$
$$v_4 = 3\lambda^2 M^6 + \lambda(15M^2 - 20M + 6)M^5,$$

kde $M=(1-\Theta)^{-1}$.Tieto výsledky nám potom umožňujú nájsť odhad parametrov rozdelenia momentovou metódou. Tak zistíme, že platí

$$\hat{\lambda} = \left(\frac{\bar{x}^3}{s^2 + \bar{x}^2}\right)^{\frac{1}{2}}, \quad \hat{\Theta} = 1 - \frac{\hat{\lambda}}{\bar{x}}.$$

Pri aplikácii na tú istú množinu dát ako v predchádzajúcich prípadoch dostávame hodnoty odhadov $\lambda=0{,}0673$ a $\Theta=0{,}1371$. Celkové výsledky sú zoradené v tabuľke číslo 4 a pre hodnotu testovacej štatistiky χ^2 -testu dobrej zhody dostávame prijateľnú hodnotu 5,919702. Táto hodnota opäť nevylučuje možnosť použiť toto rozdelenie ako model frekvencie výskytu dopravných nehôd. Pre úplnosť dodajme, že pre hodnotu G-štatistiky máme hodnotu 5,645672. Teda opäť lepší výsledok ako pri negatívnom binomickom rozdelení.

Tabuľka 4

Empirické a teoretické početnosti poistných udalostí pri použití zovšeobecneného Poissonovho rozdelenia s parametrami odhadnutými momentovou metódou.

k	n_k	Teoretické hodnoty (Parameter odhadnutý momentovou metódou)
0	3739	3739,66
1	225	219,43
2	24	32,67
3	10	6,37
4	1	1,41
5	1	0,34
≥6	0	0,09

Zdroj: Vlastný zber dát s poisťovní v roku 2004 a vlastné výpočty.

ZÁVER

Poissonovo rozdelenie, ktoré je dobrým modelom frekvencie výskytu dopravných nehôd u jednotlivého vodiča sa ukazuje ako nevhodný nástroj na modelovanie výskytu nehôd veľkej skupiny vodičov. Príčinou je predovšetkým nesplniteľnosť požiadavky homogenity celej skupiny, charakterizovaná identickou hodnotou parametra λ pre všetkých vodičov. Ako dobrý model sa ukazuje zovšeobecnené Poissonovo rozdelenie a zmesové tzv. Poissonovo-inverzné Gaussovo rozdelenie.

To predurčuje tieto rozdelenia za tzv. sčítacie rozdelenia, ktoré sú modelom výskytu udalostí pri aplikovaní kolektívneho modelu rizika. Tento model potom umožňuje pri známo sčítacom rozdelení a rozdeleniach individuálnych škôd stanoviť rozdelenie celkových škôd napr. v rámci poistného portfólia. S ohľadom na vznikajúce matematické ťažkosti sa obvykle pre výsledné rozdelenie celkových škôd odvodzujú rekurentné vzorce, ako sú napr. Panjérove rekurzie prezentované napr. v [2].

Oznam: Táto práca je podporovaná grantom VEGA 1/0931/11.

LITERATÚRA

[1] HORÁKOVÁ,G.,HUŤKA,V.: *Teória pravdepodobnosti (in Slovak)*, Ekonóm Bratislava, 2002.
[2] HORÁKOVÁ,G.,MUCHA,V.: *Teória rizika v poistení, 1.časť*, Ekonóm Bratislava, 2006.
[3] KOZUBÍK,A.: *Zložené zovšeobecnené Poissonovské rozdelenia*, 3.vedecký seminár Poistná matematika v teórii a praxi, Bratislava 2001.
[4] KOZUBÍK,A..: *Zovšeobecnenia Poissonovho rozdelenia ako modely počtu poistných udalostí*, 4.vedecký seminár Poistná matematika v teórii a praxi, Bratislava 2003.
[5] KOZUBÍK,A..: *Zovšeobecnené Poissonovo rozdelenie ako model počtu poistných udalostí v poistení motorových vozidiel*, Slovenská štatistika a demografia, 14.ročník č. 2/2004, str. 14-24.
[6] LEMAIRE,J.: *Selection Procedures of Regression Analysis Applied to Automobile Insurnce*, Mitteilungen der Schweizerischer Versicherungsmathematiker, Part I,1977, str. 143-160.
[7] LEMAIRE,J: *Automobile Insurance: Actuarial Models*, Kluwer, Boston, 1985.
[8] SOKAL,R.P. and ROHLF,F.,J: *Biometry: The Principles and Practice in Biological Research*, 3-rd Edition, Freeman, New York, 1994.