

*Michal Páleš*

## PANJEROVE REKURENTNÉ VZŤAHY V PROSTREDÍ JAZYKA R

### Úvod

Na vyjadrenie rozdelenia celkovej škody existuje v rámci aktuárskych techník množstvo postupov – numerické, aproximatívne, simulačné a rekurentné. V príspevku sa budeme teda venovať práve poslednej uvedenej metóde rekurentným vzťahom, ktoré napr. oproti využitiu konvolúcií disponujú značnou efektívnosťou výpočtu. Rekurentné vzťahy a ich využitie v aktuárskej oblasti, ich vývoj, triedy, úskalia použitia môžeme nájsť vo viacerých publikáciách. Za hlavného priekopníka v tejto oblasti považujeme kanadského aktuára H. H. Panjera (1981). V dôsledku Panjerových publikácií sa rekurentnými vzťahmi v aktuárskej oblasti začalo zaoberať mnoho ďalších autorov.

Vytváranie programových aplikácií (nakoľko numerické manuálne riešenie je značne náročné) s využitím rekurentných vzťahov, napríklad v prostredí MS Excel (VBA) tiež nie je jednoduché. Príspevok je teda zameraný na ukážku výpočtu pravdepodobnosti rozdelenia celkovej škody v prostredí jazyka R. R je voľne dostupný (open source) jazyk využívaný najmä v akademickej a vedeckej sfére. Ide o program, špecializovaný predovšetkým na štatistické výpočty, kde v štandardnej distribúcii, poprípade v podporných balíčkoch (*packages*) je implementované veľké množstvo pokročilých štatistických funkcií. Rovnako môže používateľ vytvárať vlastné programy s využitím programovacej funkcionality tohto jazyka. V kontexte s cieľom príspevku sme obmedzili vysvetľovanie základnej teórie v úvodných častiach príspevku na minimum.

### 1 PANJEROVE REKURENTNÉ VZŤAHY PRE POČET ŠKÔD

Ak počet škôd sa riadi diskretným rozdelením s nezápornými celočíselnými hodnotami patriacimi do triedy, ktorú Panjer označil  $(a, b, 0)$ . Pravdepodobnostná funkcia  $p_N(n)$  je potom vyjadrená pomocou Panjerovho vzťahu pre počet škôd

$$p_N(n) = \left( a + \frac{b}{n} \right) p_N(n-1), \quad n = 1, 2, 3, \dots \quad (1)$$

ak existujú konštanty  $a, b \in \mathbb{R}$ , pričom začiatočná hodnota pre rekurentný výpočet hodnôt náhodnej premennej  $N$  opisujúcej počet škôd je hodnota  $p_N(0)$ ,  $p_N(0) > 0$ . Jediné diskkrétne rozdelenia, nadobúdajúce nezáporné celočíselné hodnoty, ktoré vyhovujú vzťahu (1) sú Poissonovo, binomické, negatívne binomické rozdelenie a jeho špeciálny prípad geometrické rozdelenie. Napríklad konštanty  $a, b$ , ak  $N \sim \text{Po}(\lambda)$ , odvodíme na základe

porovnania dvoch po sebe idúcich hodnôt pravdepodobnostnej funkcie Poissonovho rozdelenia. A dostávame (pre dôkaz vid' [3])

$$a = 0, b = \lambda$$

## 2 PANJEROVE REKURENTNÉ VZŤAHY PRE CELKOVÚ ŠKODU

Panjerov rekurentný vzťah, vyjadrujúci zložené rozdelenie celkovej škody, možno využiť, ak počet škôd sa riadi diskretným rozdelením, ktoré patrí do jednotlivých tried vyhovujúcich Panjerovmu rekurentnému vzťahu, ktoré sme uviedli v časti 1 a individuálna výška škody je tiež diskretná náhodná premenná (označme  $X$ ) s pravdepodobnostnou funkciou  $p_X(x)$ .

Ak pre náhodnú premennú popisujúcu celkovú škodu  $S$ , platí

$$S = \sum_{i=1}^N X_i \quad (2)$$

potom vzťah pre pravdepodobnostnú funkciu rozdelenia celkovej škody  $p_S(x)$  má tvar (podľa [3])

$$p_S(x) = \frac{1}{1 - \alpha \cdot p_X(0)} \left( \sum_{k=1}^x \left( \alpha + \frac{b \cdot k}{x} \right) p_X(k) \cdot p_S(x - k) \right) \quad (3)$$

Vzťah (3) je Panjerov rekurentný vzťah na výpočet funkčných hodnôt pravdepodobnostnej funkcie zloženého rozdelenia s počtom škôd vyhovujúcim rozdeleniam patriacich do triedy  $(a, b, 0)$  s počiatočnou podmienkou

$$p_S(0) = p_N(0) + \sum_{n=1}^{\infty} p_N(n) \cdot p_X^{*n}(0) \quad (4)$$

Ak  $N \sim Po(l)$  potom vzťah (3) má tvar

$$p_S(x) = \sum_{k=1}^x \left( \frac{\lambda \cdot k}{x} \right) \cdot p_X(k) \cdot p_S(x - k) \quad (5)$$

a pre počiatočnú podmienku platí

$$p_S(0) = e^{-\lambda} \quad (6)$$

### 3 PRAKTICKÁ UKÁŽKA VYUŽITIA PANJEROVÝCH REKURENTNÝCH VZŤAHOV

Uvažujme situáciu, že počet škôd sa riadi Poissonovým rozdelením  $N \sim Po(4)$  a výška individuálnej výšky škody (v modelových jednotkách) je daná pravdepodobnostnou tabuľkou

$x$	1	2	3	$\Sigma$
$p_X(x)$	0,25	0,50	0,25	1

Určme zložené Poissonovo rozdelenie  $S \sim CoPo(4; p_X(x))$  s využitím vzťahu (5).

Ak  $N \sim Po(4)$  potom pre konštanty  $a$  a  $b$  platí

$$a = 0, \quad b = \lambda = 4$$

Potom pre prvé štyri hodnoty pravdepodobnostnej funkcie rozdelenia celkovej škody

$p_S(x)$  platí

$$p_S(0) = e^{-4} = 0,01832$$

$$p_S(1) = 4 \cdot p_X(1) \cdot p_S(0) = 4 \cdot \frac{1}{4} \cdot 0,01832 = 0,01832$$

$$\begin{aligned} p_S(2) &= \sum_{k=1}^2 \binom{4 \cdot k}{2} \cdot p_X(k) \cdot p_S(2-k) = \\ &= \frac{4}{2} \cdot p_X(1) \cdot p_S(1) + \frac{8}{2} \cdot p_X(2) \cdot p_S(0) = \\ &= \frac{4}{2} \cdot \frac{1}{4} \cdot 0,01832 + \frac{8}{2} \cdot \frac{1}{2} \cdot 0,01832 = \\ &= \frac{1}{2} \cdot 0,01832 + 2 \cdot 0,01832 = 0,04579 \end{aligned}$$

$$p_S(3) = \sum_{k=1}^3 \binom{4 \cdot k}{3} \cdot p_X(k) \cdot p_S(3-k) = \dots = 0,05799$$

a tak ďalej.

## 4 REALIZÁCIA VÝPOČTU V JAZYKU R

Na výpočet sumy vo vzťahu (5), by sme mohli použiť cyklus, ale pretože R je interpretovaný jazyk, je vhodné „vektorizovať“ výpočty a tak nahradiť cyklus funkciou sum. Pri numerickom výpočte, podľa vzťahu (5), sme si mohli všimnúť opakujúce sa hodnoty  $\lambda/x$ , ktoré vlastne ovplyvňujú tri vektory

$$\begin{aligned} & (1, \dots, m) \\ & (p_x(1), \dots, p_x(m)) \\ & (p_s(x-1), \dots, p_s(x-m)) \end{aligned}$$

kde  $m = \min\{x, r\}$ , pričom  $r$  je maximálny index pre ktorý  $p_r > 0$ . Program pre využitie Panjerových rekurentných vzťahov v situácii, ktorá je definovaná príkladom uvedeným v predchádzajúcej časti príspevku, môžeme písať nasledovne

```
Panjer.Poisson <- function (p, lambda)
{ if(sum(p)>1||any(p<0)) stop("p parameter not a density")
  if(lambda * sum(p) > 727) stop ("Underflow")
  cumul <- f <- exp(-lambda * sum(p))
  r <- length(p)
  x <- 0
  repeat
  { x <- x+1
    m <- min(x, r)
    last<-lambda / x * sum(1:m * head(p,m) * rev(tail(f,m)))
    f <- c(f,last)
    cumul <- cumul + last
    if (cumul > 0.99999999) break  }
  return(f) }
Panjer.Poisson(c(0.25,0.5,0.25), 4)
pp <- Panjer.Poisson(c(0.25,0.5,0.25), 4)
```

Pri písaní programu sme ponechali ekvivalenty pri definovaní objektov a pomocných textov v anglickom jazyku. Z dôvodu označenia pravdepodobnostných funkcií  $p_x(x)$ ,  $p_s(x)$  a následnej nejednoznačnosti bez uvedenia dolných indexov, sme tieto označili ako  $p, f$ .

Syntax programu vypíše preddefinované chybové hlásenie v prípade ak

$$\sum p_x(x) > 1 \text{ resp. } p_x(x) < 0$$

Označenie  $\lambda \equiv \lambda$ , pričom hodnota parametra  $\lambda$  nemôže byť príliš vysoká. Výpočtový problém (označovaný v angl. ako *underflow*; „pretečenie“) v oblasti stanovenia hodnoty parametra  $\lambda$  nastáva (v štandardných systémoch MS Windows), ak

$$\lambda(1 - p_X(0)) > 727$$

nakolko hodnota  $p_S(0)$  je potom príliš malá. R využíva pri výpočtoch 64-bit presnosť, ale pri systémoch s 80-bit možno uvažovať pri analýze portfólia aj s parametrom  $\lambda = 11340$ , kde počiatočná hodnota  $p_S(0) \cong 10^{-5000}$ . V starších programovacích jazykoch (48-bit) tento problém už vznikal, ak  $\lambda(1 - p_X(0)) \geq 88$ . Napríklad pre  $n$  rozmerné portfólia životného poistenia (ak  $p_N(0) = 0,005$ ) dochádza k tomu javu, ak  $n = 17600$ . S týmto problémom sa možno vysporiadať využitím kompilovaného externého kódu, ktorý využíva rozšírenie pre spresnenie výpočtov. Tento spôsob môže napríklad pomôcť eliminovať numerickú nestabilitu pri zložennom binomickom rozdelení (viac vid' problematika nestability rekurentných vzťahov zloženého binomického rozdelenia).

Vstupné hodnoty (pravdepodobnosti  $p_X(x)$  a parameter  $\lambda$ ) používateľ zadáva ako argument funkcie `Panjer.Poisson(...)`. Objekt `pp` potom definuje výsledok riešenia a teda predstavuje vektor, ktorého zložky sú jednotlivé hodnoty pravdepodobnostnej funkcie  $p_S(x)$ , teda

$$\bar{p}_S = (p_S(0), p_S(1), \dots, p_S(w))$$

pričom pre  $w$  platí

$$p_S(0) + p_S(1) + \dots + p_S(w) > 1 - 10^{-8}$$

Pre výstup konkrétnej hodnoty  $x$  môžeme využiť zápis `pp[x]`, pričom platí `pp[1] = p_S(0)`, a tak ďalej. Potom pre hodnoty `pp[1]` až `pp[4]` dostávame výsledok identický s výpočtom v časti 3, a to

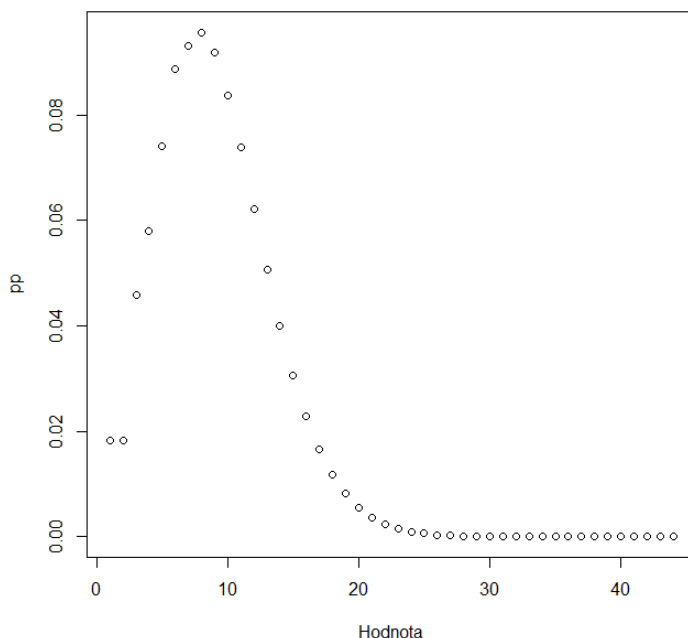
$$0.01831564, \quad 0.01831564, \quad 0.04578910, \quad 0.05799952$$

Je zrejmé, že výsledkom funkcie `sum(pp)` je hodnota 1. Graf rozdelenia pravdepodobnosti celkovej škody dostávame po zadaní príkazu `plot(pp, xlab="Hodnota")`, vid' obrázok 1.

Výpočet hodnôt pravdepodobnosti zloženého rozdelenia celkového počtu škôd možno komparatívne určiť aj pomocou iných metód – napríklad pomocou rýchlej Fourierovej transformácie. Na výpočet je možné rovnako využiť prostredie jazyka R, ktorý tu disponuje zabudovanou funkciou `fft`. Výpočet s využitím rýchlej Fourierovej transformácie sa tým stáva v mnohých prípadoch jednoduchší a celkovo dynamickejší ako

pri využití Panjerových rekurentných vzťahov. Problematika výpočtu si však vyžaduje nastudovanie pokročilého teoretického aparátu.

Obrázok 1: Graf pravdepodobnostnej funkcie  $p_s(x)$



## Záver

Panjerov rekurentný vzťah sa využíva najmä v prípade ak počet škôd sa štandardne riadi vyhovujúcim diskretným rozdelením triedy, pričom náhodná premenná individuálna výška škody je diskretná. Zvyčajne je v tomto prípade určená pravdepodobnostnou tabuľkou bez špecifikácie rozdelenia tzn. ku každej hodnote individuálnej výšky škody je priradená pravdepodobnosť. V prípade, že výška škody sa riadi konkrétnym spojitým rozdelením pravdepodobnosti s príslušnými parametrami využíva sa napríklad metóda diskretizácie spojitý individuálnej výšky škody. Výpočet hodnôt pravdepodobnostnej funkcie rozdelenia celkovej škody pomocou Panjerových vzťahov je veľmi účinné a jeho význam v aktuárskych výpočtoch ho stavia na popredné miesto (pozornosť je potrebné venovať tiež stabilite týchto algoritmov). Metóda výpočtu, aj napriek tomu, že má výhodu, že netreba priamo počítať konvolúcie, si vyžaduje počítačovú realizáciu. Tu sme využili a popísali jazyk R ako účinný nástroj pre túto realizáciu a rovnako sme uviedli aj viaceré úskalia, ktoré môžu pri programovaní resp. výpočte nastať.

## Kľúčové slová

jazyk R, aktuárske analýzy, teória rizika v poistení, rekurentné vzťahy, Panjerove rekurentné vzťahy

**Klasifikácia JEL**

G22; C63

**LITERATÚRA**

- [1] KAAS, R. – GOOVAERTS, M. – DHAENE, J. – DENUIT, M. 2008. *Modern actuarial risk theory using R*. Berlin : Springer, 2008. ISBN 978-3-642-03407-7.
- [2] KLUGMAN, S. A. – PANJER, H. H. – WILLMOT, G. E. 2009. *Loss Models (From Data to Decision)*. New York : John Wiley & Sons, 2009. ISBN 978-0-470-48744-0.
- [3] PÁLEŠ, M. 2012. *Rekurentné vzťahy pre aktuárov a ich aplikácia v oblasti zaistenia* : dizertačná práca. Bratislava : Ekonomická univerzita v Bratislave, 2012.
- [4] PÁLEŠ, M. 2010. Teória rizika a jej aplikácia v rozhodovacích procesoch komerčných poisťovní. In *IMEA 2010* : 10th international conference of postgraduate students and young scientists in informatics, management economics and administration, Seč, Chrudim 26th-27th April, 2010. Pardubice : Univerzita Pardubice, 2010. ISBN 978-80-7395-254-9.

**RESUMÉ**

Matematicko-štatistické modely sú dôležitým nástrojom analýzy poisťných rizík a významne podporujú všetky procesy v tvorbe vývoja produktov v komerčných poisťovniach. S rámcom smernice Solvency II sa teória rizika stáva dôležitým nástrojom pre tvorbu aktuárskych modelov, ktoré sú oporným prostriedkom pre manažment poisťovní. Zavádzanie rôznych modelovacích techník otvára široké možnosti aplikácie postupov a metód teórie rizika, a to najmä využitie pravdepodobnostných rozdelení v poisťovníctve individuálnych poisťných plnení a následne celkovej škody, v súvislosti s tvorbou poisťných modelov s podporou výpočtovej techniky. Príspevok prezentuje programovú aplikáciu v jazyku R, ktorá môže byť využitá pre výpočet hodnôt pravdepodobnostnej funkcie rozdelenia celkovej škody, ak počet škôd má napr. Poissonovo rozdelenie s daným parametrom a individuálna výška škody je diskretná daná pravdepodobnostnou tabuľkou, s využitím Panjerových rekurentných vzťahov. Uviedli sme odvodené Panjerové rekurentné vzťahy pre počet škôd aj celkovú škodu danej triedy. Tieto vzťahy aplikujeme najskôr pre ukážku manuálneho výpočtu a následne pomocou napísaného programu v jazyku R. Rovnako sa zaoberáme výpočtovými obmedzeniami, ktoré sú s touto problematikou spojené. Vytvorenie kvalitného a spoľahlivého aktuárskeho modelu (scenáru), opierajúceho sa o pevný matematicko-štatistický základ, môže poisťovni zabezpečiť značnú konkurenčnú výhodu pri tvorbe nového produktu. Nájst' však presný model rozdelenia pravdepodobnosti celkových škôd pri rôznych druhoch poistenia je často komplexná a náročná úloha. Preto je dôležité snažiť sa hľadať jednoduchšie metódy a v čo najväčšej miere zapájať do tohto systému výpočtovú techniku. Jedným z takýchto zjednodušení, ako sme ukázali, je aplikácia Panjerových rekurentných vzťahov a jazyka R.

## SUMMARY

Mathematical-statistical models are important tools for analysis of insurance risk. With Solvency II project, the risk theory becomes an important tool for the creation of insurance models. Implementation of various modelling techniques opens up wide possibilities of application procedures and methods of risk theory, mainly use of probability distribution in the individual risk model and then collective risk model in connection with the generation of insurance models to support computer technology. The aim of paper is presentation of application in R language. It can be used to calculate a probability distribution of total claim consider a compound Poisson distribution with parameter lambda and individual claim is discrete with implementing Panjer's recursion. We introduced derived relations for the number of claims and the total claims. Recursions apply for manually calculation at first and then for calculation with written program in language R. Further discuss about computational constraints that are associated with Panjer's recursion. Creating a significant and reliable actuarial model (scenario), based on a solid mathematical and statistical apparatus, can be providing a significant competitive advantage in the development of new product for insurance company. Find the exact model of the probability distribution of the total claims in different types of insurance is complex and challenging task. It is important to apply easier methods and use PC systems. We showed the application and connection Panjer's recursion and language R.

## **Kontakt**

Ing. Michal Páleš, PhD., Katedra matematiky a aktuárstva, Fakulta hospodárskej informatiky Ekonomická univerzita v Bratislave, Dolnozemska cesta 1/b, 852 35 Bratislava, tel.: +421 2/672 95 841, e-mail: [pales.euba@gmail.com](mailto:pales.euba@gmail.com)