UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

KONŠTRUKCIA A VALIDÁCIA MATEMATICKÝCH MODELOV PRE OBLASŤ AKTUÁRSTVA

Diplomová práca

2023

Bc. Zuzana Žuffová

UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

KONŠTRUKCIA A VALIDÁCIA MATEMATICKÝCH MODELOV PRE OBLASŤ AKTUÁRSTVA

Diplomová práca

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	Matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci:	Mgr. Gábor Szűcs, PhD.

Bratislava, 2023

Bc. Zuzana Žuffová

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Zuzana Žuffová
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Konštrukcia a validácia matematických modelov pre oblasť aktuárstva
Construction and validation of mathematical models for actuarial science

Anotácia: Matematické modelovanie je neodmysliteľnou súčasťou aktuárskej vedy, v ktorej sa spájajú techniky z teórie pravdepodobnosti, matematickej štatistiky, finančnej matematiky, demografickej štatistiky či dátovej vedy. Jednou z hlavných úloh aktuára je, aby na základe minulých údajov a dostupných informácií zostrojil dostatočne presný, efektívny a robustný matematický model, pomocou ktorého potom dokáže analyzovať riziká, vytvárať predikcie a odhadovať budúci vývoj skúmaných veličín.

Vedúci: Mgr. Gábor Szűcs, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Marek Fila, DrSc.

Dátum zadania: 11.01.2022

Dátum schválenia: 14.01.2022 prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Abstrakt

Matematické modelovanie v oblasti aktuárstva predstavuje dôležitý nástroj pre posudzovanie a riadenie finančného rizika. Použitím techník z teórie pravdepodobnosti, matematickej štatistiky či demografickej štatistiky je možné navrhnúť funkčné a efektívne matematické modely, ktoré dostatočne presne opisujú skutočný svet. Táto záverečná práca poskytuje prehľadné a zrozumiteľné spracovanie postupu tvorby aktuárskych modelov. Pozornosť je venovaná aktuárskemu modelovaniu v modeli individuálneho rizika a rovnako aktuárskemu modelovaniu v teórii ruinovania. Problematika je vysvetľovaná na ilustratívnych príkladoch, ktorých základom sú umelé dáta, generované v prostredí štatistického softvéru R. Kalibrácia modelov, validácia modelov a všetky ostatné výpočty sú taktiež realizované v tomto softvéri. Cieľom diplomovej práce je uviesť vybrané okruhy matematického modelovania do kontextu tvorby aktuárskych modelov.

Kľúčové slová: aktuárske modelovanie, model individuálneho rizika, teória ruinovania.

Abstract

Mathematical modeling in actuarial science is an important tool for assessing and managing financial risk. By using techniques from the theory of probability, mathematical statistics, or demographic statistics, it is possible to design functional and effective mathematical models that describe the real world accurately enough. This final thesis provides a clear and comprehensible treatment of the process of creating actuarial models. Attention is paid to actuarial modeling in the individual risk model and actuarial modeling in the ruin theory. The issue is explained using illustrative examples based on artificial data generated in the statistical software R. Model calibration, model validation and all other calculations are also implemented in this software. The aim of the thesis is to introduce selected areas of mathematical modeling into the context of the creation of actuarial models.

Keywords: actuarial modeling, individual risk model, ruin theory.

Predhovor

Aktuárske modely, ako aj mnohé iné modely, nedokážu opísať svet dokonale presne, no sú v mnohých smeroch užitočné. S aktuárskymi modelmi sa môžeme stretnúť v rôznych oblastiach matematického modelovania. Moje prvé stretnutie sa s demografickým modelovaním na predmete *Demografická štatistika* ma zaujalo natoľko, že som sa rozhodla rozšíriť svoje vedomosti aj v iných tematických okruhoch matematického modelovania v oblasti aktuárstva. Môžem tvrdiť, že písaním tejto diplomovej práce som nadobudla mnoho nových znalostí, a to najmä z odvetvia poisťovníctva, aktuárskej vedy, teórie pravdepodobnosti a matematickej štatistiky. Taktiež usudzujem, že sa mi podarilo v dostačujúcej miere splniť vopred stanovené ciele diplomovej práce. Nakoniec verím, že výsledky tejto záverečnej práce môžu byť inšpiráciou pre tvorbu modelov aj v ďalších okruhoch aktuárskeho modelovania.

Na tomto mieste by som sa zároveň veľmi rada poďakovala vedúcemu diplomovej práce Mgr. Gáborovi Szűcsovi, PhD., za všetky cenné pripomienky a odporúčania, za jeho ochotu pomôcť a nadmieru ľudský prístup. Rovnako ďakujem aj svojim najbližším, ktorí ma podporovali počas celého štúdia a aj pri písaní tejto diplomovej práce.

Obsah

Úvod

Formovanie aktuárskej vedy ako vedeckej disciplíny siaha až do konca 17. storočia, kde začiatky jej vzniku sú úzko späté s rozvojom demografie a so zvýšeným záujmom o zhodnocovanie úspor. V tomto období boli vytvorené prvé úmrtnostné tabuľky a vznikali prvé životné poisťovne, ktorých výkonný riaditeľ bol označovaný ako aktuár [12]. Momentálne si pod pojmom aktuár predstavíme odborníka v oblasti poistnej matematiky, ktorý vytvára produkty životného či neživotného poistenia, odhaduje výšku technických rezerv a zodpovedá za správnosť výpočtov. Aktuárska veda v sebe spája niekoľko vzájomne súvisiacich predmetov vrátane matematiky, teórie pravdepodobnosti, matematickej štatistiky, finančnej matematiky či demografickej štatistiky. Dnes máme k dispozícii mnoho matematických modelov, vďaka ktorým sme schopní čo najpresnejšie ohodnotiť riziká súvisiace s poistením, odhadnúť výšku či počet poistných nárokov, výšku technických rezerv a podobne, čo je nesmierne dôležité pre správne fungovanie sektoru poistenia a zaistenia.

Tvorba aktuárskych modelov nám poskytuje širokú oblasť výberu tematických okruhov matematického modelovania: od demografického modelovania cez finančnomatematické modelovanie výnosových kriviek až po matematické modelovanie v modeli kolektívneho či individuálneho rizika a mnoho ďalších. V tejto diplomovej práci sa zaoberáme konkrétne aktuárskym modelovaním v modeli individuálneho rizika a aktuárskym modelovaním v teórii ruinovania, pričom dôraz kladieme na postupnosť technických krokov používaných pri tvorbe aktuárskych modelov.

Cieľom diplomovej práce je prehľadne spracovať a priblížiť čitateľovi postup tvorby aktuárskych modelov a oboznámiť ho s aktuárskymi metodikami používanými pri konštruovaní a validácii aktuárskych modelov. Charakter diplomovej práce je miestami pedagogický, metodiky sú demonštrované na ilustratívnych príkladoch, a tak môže diplomová práca v budúcnosti slúžiť aj ako učebný text.

Diplomová práca je rozdelená do piatich kapitol, kde prvá kapitola predstavuje všeobecný úvod do problematiky diplomovej práce, a rovnako prináša prehľad existujúcej literatúry. Druhá kapitola sa venuje vybraným matematickým, pravdepodobnostným a štatistickým metodikám používaných pri konštruovaní a validácii aktuárskych modelov, kde zároveň definujeme matematické označenia či skratky používané v ďal-

ších častiach diplomovej práce. V tretej kapitole si predstavíme okruhy matematického modelovania v oblasti aktuárstva, a tiež postupnosť krokov používaných pri tvorbe aktuárskych modelov. Najrozsiahlejšia je štvrtá a piata kapitola, kde si najskôr priblížime model individuálneho rizika, a následne prezentujeme postup aktuárskeho modelovania v modeli individuálneho rizika na ilustratívnom príklade. V piatej kapitole sa zameriame na teóriu ruinovania a postupnosť krokov aktuárskeho modelovania v teórii ruinovania opäť demonštrujeme na ilustratívnom príklade. Pri znázornení niektorých výsledkov z posledných dvoch kapitol používame aj grafické ilustrácie.

1 Všeobecný prehľad z oblasti aktuárskeho modelovania

Pod slovným spojením aktuárska veda máme na mysli disciplínu, ktorá sa zaoberá matematickým modelovaním a analýzou rizika, a to najmä v oblasti poistenia a financií, pričom matematický základ spočíva v teórii pravdepodobnosti a štatistických metódach. Aktuárske modelovanie či inými slovami poistno-matematické modelovanie je kľúčovým prvkom aktuárskej vedy. Aktuárske modely sa používajú na odhadovanie pravdepodobnosti vzniku náhodných udalostí či na určenie výšky kapitálu, ktorý by sa mal vyhradiť na pokrytie potenciálnych strát. Modelovanie v oblasti aktuárstva vyžaduje znalosť rozličných nástrojov teórie pravdepodobnosti vrátane známych rozdelení pravdepodobnosti, centrálnej limitnej vety a podobne. Taktiež sa využívajú rôzne štatistické metódy či techniky, ako napríklad testy dobrej zhody alebo Monte Carlo simulácie.

Existuje mnoho oblastí aktuárskeho modelovania, pričom ako sme vyššie naznačili, aktuárske modely sa používajú predovšetkým v poisťovníctve, a to v životnom, neživotnom i zdravotnom poistení. V prípade životného poistenia je možné využiť znalosti demografického modelovania a zamerať sa na modelovanie úmrtnostných kriviek pre presnejší odhad pravdepodobnosti úmrtia a súvisiacich finančných výdavkov. Aktuárske modelovanie je možné nájsť aj v oblasti neživotného poistenia, či už pri modelovaní výšky alebo počtu poistných nárokov napríklad v modeli kolektívneho rizika, čo vedie k odhadovaniu celkovej výšky budúcich výdavkov. Nakoniec aj v oblasti zdravotného poistenia môžeme použiť údaje o výškach poistných nárokov za zdravotnú starostlivosť, a odhadovať tak výšku budúcich výdavkov pri plánovaní financovania zdravotnej starostlivosti v rámci systému zdravotného poistenia.

V konečnom dôsledku aktuárske modelovanie predstavuje nástroj pomáhania jednotlivcom a organizáciám robiť správne finančné rozhodnutia. Je preto potrebné vzdelávať mladých aktuárov, a to v rôznych oblastiach modelovania. Motivovaní touto myšlienkou sme sa rozhodli predstaviť v tejto diplomovej práci jednotlivé kroky postupu tvorby aktuárskych modelov spolu s ilustráciou ich použitia v niektorých konkrétnych tematických okruhoch modelovania, čo predstavuje možnú pridanú hodnotu tejto

práce. Zároveň sú v pasívnej prílohe diplomovej práce priložené aj programové kódy k neskôr prezentovaným ilustratívnym príkladom, pri ktorých tvorbe vychádzame z viacerých známych kníh. Podkladom pre štúdium teórie pravdepodobnosti a štatistiky pri tvorbe diplomovej práce sú knihy: *Základy matematické statistiky* od Jiřího Anděla [1], *Teorie pravděpodobnosti* od Alfréda Rényiho [11] a kniha rovnako s názvom *Teória pravdepodobnosti* od Jeleny Sergejevny Ventceľovej [15]. Pri prvej konkrétnej aplikácii aktuárskeho modelovania v modeli individuálneho rizika je najskôr potrebné oboznámiť sa s týmto modelom. Model individuálneho rizika je opísaný vo viacerých knihách, z ktorých vychádzame aj my. Ide napríklad o knihu od Davida C. M. Dicksona s názvom *Insurance Risk and Ruin* [3] alebo o knihu od Roba Kaasa a kolektívu autorov s názvom *Modern Actuarial Risk Theory* [6]. V druhom prípade, pri popise aktuárskeho modelovania v teórii ruinovania, využívame pre štúdium tejto teórie opäť dve vyššie spomenuté publikácie od D. C. M. Dicksona a R. Kaasa a kol., a tiež knihu od Galiny Horákovej a kol. s názvom *Teória rizika v poistení* [5] a knihu od Thomasa Mikoscha s názvom *Non-Life Insurance Mathematics* [8]. Vyššie uvedené publikácie nám tak poskytujú všeobecné základy pre model individuálneho rizika aj pre teóriu ruinovania, no tieto knihy nie sú písané v štýle a kontexte tvorby aktuárskych modelov. Celou diplomovou prácou nás sprevádzajú aj hodnotné skriptá od Gábora Szűcsa, prvé s názvom *Matematické modelovanie v poisťovníctve* [12] a druhé s názvom *Pravdepodobnostné modely v poisťovníctve* [13].

2 Štatistické metódy v oblasti aktuárstva

Aktuárske modelovanie sa zakladá na určitých matematických, pravdepodobnostných a štatistických znalostiach, ktoré si v tejto kapitole priblížime. V podkapitole 2.1 uvedieme základné pojmy a definície z oblasti pravdepodobnosti a matematickej štatistiky, ktoré sa budú v ďalších kapitolách často vyskytovať. Vychádzame pritom predovšetkým z kníh [1] a [15]. Následne v podkapitolách 2.2 a 2.3 definujeme vybrané rozdelenia pravdepodobnosti, spolu s ich matematickými označeniami a skratkami, pričom informácie čerpáme prevažne z publikácií [1] a [12]. Nakoniec sa v podkapitole 2.4 zameriame na niektoré testy dobrej zhody, ktoré budeme v ďalších kapitolách používať pri konštruovaní a validácii aktuárskych modelov.

2.1 Základné pojmy a definície

Uveďme (bez nároku na úplnosť) niekoľko dôležitých pojmov z teórie pravdepodobnosti a matematickej štatistiky. Jedným z najvýznamnejších pojmov teórie pravdepodobnosti je pojem náhodnej premennej, pričom je však potrebné oboznámiť sa najskôr s pojmom náhodný experiment. Za náhodný experiment považujeme takú situáciu, v ktorej výsledky pri viacerých opakovaniach môžu nadobúdať za rovnakých počiatočných podmienok rozdielne hodnoty. Potom podľa [15] je **náhodná premenná** taká premenná, ktorá výsledkom náhodného experimentu nadobúda nejakú hodnotu, pričom je vopred neznáme, akú konkrétne. Náhodnú premennú je tak možné definovať aj ako funkciu náhodného experimentu. Náhodné premenné sa zvyknú označovať veľkými písmenami, zväčša ako X, a delíme ich na dva typy, a to na diskrétne náhodné premenné a spojité náhodné premenné. Medzi diskrétne náhodné premenné patria tie premenné, ktoré môžu nadobudnúť iba konečný počet reálnych hodnôt, ktoré je možné vopred vymenovať, a naopak, spojité náhodné premenné nadobúdajú nekonečný počet reálnych hodnôt, ktoré spojite vyplňujú nejaký interval. Ďalším dôležitým pojmom, ktorý je vhodné uviesť, je pojem distribučná funkcia. **Distribučná funkcia** náhodnej premennej X je reálna funkcia definovaná predpisom:

$$F(x) = \mathrm{P}(X < x), \tag{2.1}$$

kde P označuje takzvanú funkciu pravdepodobnosti. Distribučná funkcia je najvšeobec-

nejšou charakteristikou náhodnej premennej, a zároveň jednoznačne určuje rozdelenie pravdepodobnosti náhodnej premennej. Dá sa tiež ukázať, že distribučná funkcia je neklesajúca, zľava spojitá funkcia pre ktorú platí:

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{a} \quad \lim_{x\to\infty} F(x) = 1. \tag{2.2}$$

Distribučnú funkciu je možné definovať pre diskrétne i spojité náhodné premenné. V prípade diskrétnych rozdelení sa distribučná funkcia nazýva aj skoková funkcia a obsahuje spočítateľne veľa bodov nespojitosti, ktoré môžeme označiť ako $x_1, x_2, \ldots, x_k, \ldots$ a k ním prislúchajúcu veľkosť skokov ako $p_1, p_2, \ldots, p_k, \ldots$ Potom sa dá ukázať, že $p_k = \mathrm{P}(X = x_k)$. Zovšeobecnením tohto vzťahu dostávame **pravdepodobnostnú funkciu** definovanú ako $p(x) = \mathrm{P}(X = x)$. Pre distribučnú funkciu diskrétnych rozdelení pravdepodobnosti sa potom zvykne používať aj zápis $\mathrm{P}(X \leq x)$ a zápis (2.1) používame skôr pre spojité rozdelenia pravdepodobnosti. Je teda potrebné venovať pozornosť použitiu týchto znamienok, pretože môžu spôsobiť rozdiel v prípade diskrétnych rozdelení. V prípade spojitých rozdelení pravdepodobnosti je možné distribučnú funkciu definovať predpisom:

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t, \tag{2.3}$$

kde funkcia $f(t)$ sa nazýva **hustota rozdelenia pravdepodobnosti** (v našej práci ju pre jednoduchosť nazývame len ako *hustota*). Platí teda vzťah $F'(x) = f(x)$. Keďže distribučná funkcia je neklesajúca, potom hustota rozdelenia je nezáporná a tiež je známe, že celkový obsah plochy pod funkciou hustoty je rovný 1. Nakoniec definujme ešte momentovú vytvárajúcu funkciu náhodnej premennej. **Momentová vytvárajúca funkcia** náhodnej premennej X označovaná ako $M(t)$ je definovaná nasledovne:

$$M(t) = \mathrm{E}\left(\mathrm{e}^{tX}\right), \quad \text{pre } t \in \mathbb{R}, \tag{2.4}$$

pričom momentovú vytvárajúcu funkciu pre diskrétnu náhodnú premennú X je možné počítať ako:

$$M(t) = \sum_{i=1}^{\infty} \mathrm{e}^{tx_i}\mathrm{P}(X = x_i), \tag{2.5}$$

zatiaľ čo pre spojitú náhodnú premennú X počítame momentovú vytvárajúcu funkciu nasledujúcim spôsobom:

$$M(t) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x)\mathrm{d}x, \tag{2.6}$$

a to len za predpokladu, že postupnosť $\{e^{tx_i}P(X = x_i)\}_{i=1}^{\infty}$ je konvergentná a integrál vo vzťahu (2.6) existuje [12].

2.2 Spojité rozdelenia pravdepodobnosti

Oboznámme sa najskôr s niektorými známymi spojitými rozdeleniami pravdepodobnosti, pri ktorých definícii vychádzame prevažne z publikácii [1] a [12]. K vybraným rozdeleniam pravdepodobnosti navyše uvádzame aj súvisiace funkcie implementované v štatistickom softvéri R [9].

Normálne rozdelenie pravdepodobnosti

Náhodná premenná X sa riadi normálnym rozdelením pravdepodobnosti s parametrami $\mu \in \mathbb{R}$ a $\sigma \in \mathbb{R}^+$, ak má hustotu [1]:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad \text{pre } x \in \mathbb{R}.$$

Normálne rozdelenie pravdepodobnosti označujeme symbolom $\mathcal{N}(\mu, \sigma^2)$ a hovorí sa mu aj Gaussovo rozdelenie pravdepodobnosti. V prostredí štatistického softvéru R je normálne rozdelenie označované skratkou `norm` a spája sa s ním aj niekoľko známych a užitočných funkcií, a to napríklad funkcia `pnorm` pre distribučnú funkciu normálneho rozdelenia, `dnorm` pre funkciu hustoty normálneho rozdelenia, `qnorm` pre kvantilovú funkciu normálneho rozdelenia či `rnorm` pre generovanie pseudonáhodných čísel z normálneho rozdelenia pravdepodobnosti.

Log-normálne rozdelenie pravdepodobnosti

Náhodná premenná X sa riadi log-normálnym rozdelením pravdepodobnosti s parametrami $\mu \in \mathbb{R}$ a $\sigma \in \mathbb{R}^+$, ak má hustotu [12]:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}, \quad \text{pre } x > 0.$$

Log-normálne rozdelenie pravdepodobnosti označujeme symbolom $\mathrm{LN}(\mu, \sigma^2)$, pričom parameter μ sa nazýva *log-scale* parameter a parameter σ *log-shape* parameter. Log-normálne rozdelenie pravdepodobnosti reprezentuje v softvéri R skratka `lnorm` a sú v ňom tiež zabudované funkcie `plnorm`, `dlnorm`, `qlnorm` a `rlnorm` pre distribučnú funkciu, hustotu, kvantilovú funkciu a generovanie realizácii log-normálneho rozdelenia.

Exponenciálne rozdelenie pravdepodobnosti

Náhodná premenná X sa riadi exponenciálnym rozdelením pravdepodobnosti s parametrom $\lambda \in \mathbb{R}^+$, ak má hustotu [12]:

$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \quad \text{pre } x \geq 0.$$

Exponenciálne rozdelenie pravdepodobnosti označujeme symbolom $\mathrm{Exp}(\lambda)$, kde parameter λ sa nazýva *rate* parameter. V prostredí štatistického softvéru R označujeme exponenciálne rozdelenie pravdepodobnosti skratkou `exp`. Tiež sú dostupné funkcie `pexp`, `dexp` a `qexp`, vďaka ktorým vieme jednoducho vypočítať konkrétne hodnoty distribučnej funkcie, funkcie hustoty či kvantilovej funkcie exponenciálneho rozdelenia pravdepodobnosti. Pre generovanie pseudonáhodných čísel z exponenciálneho rozdelenia môžeme použiť funkciu `rexp`.

Gama rozdelenie pravdepodobnosti

Náhodná premenná X sa riadi gama rozdelením pravdepodobnosti s parametrami $\alpha \in \mathbb{R}^+$ a $\beta \in \mathbb{R}^+$, ak má hustotu [1], [12]:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \mathrm{e}^{-\beta x} x^{\alpha-1}, \quad \text{pre } x \geq 0,$$

kde $\Gamma(\alpha)$ predstavuje gama funkciu definovanú vzťahom:

$$\Gamma(\alpha) = \int_0^\infty \mathrm{e}^{-x} x^{\alpha-1} \mathrm{d}x.$$

Gama rozdelenie pravdepodobnosti označujeme symbolom $\mathrm{Gama}(\alpha, \beta)$, pričom parameter α sa nazýva *shape* parameter a parameter β je *rate* parameter. Navyše platí, že exponenciálne rozdelenie pravdepodobnosti je špeciálnym prípadom gama rozdelenia pravdepodobnosti, ktoré dostaneme, ak *shape* parameter gama rozdelenia pravdepodobnosti $\alpha = 1$. Gama rozdelenie pravdepodobnosti reprezentuje v softvéri R skratka `gamma`. Hodnoty distribučnej funkcie, funkcie hustoty a kvantilovej funkcie je možné vypočítať pomocou funkcií `pgamma`, `dgamma` a `qgamma`, a tiež je možné generovať pseudonáhodné realizácie gama rozdelenia pomocou funkcie `rgamma`.

Weibullovo rozdelenie pravdepodobnosti

Náhodná premenná X sa riadi Weibullovým rozdelením pravdepodobnosti s parametrami $\kappa \in \mathbb{R}^+$ a $\theta \in \mathbb{R}^+$, ak má hustotu [12]:

$$f(x) = \frac{\kappa x^{\kappa-1}}{\theta^\kappa} \exp\left\{-\left(\frac{x}{\theta}\right)^\kappa\right\}, \quad \text{pre } x > 0.$$

Weibullovo rozdelenie pravdepodobnosti označujeme symbolom $\text{Wei}(\kappa, \theta)$, pričom parameter κ sa nazýva *shape* parameter a parameter θ *scale* parameter. Taktiež platí, že Weibullovo rozdelenie pravdepodobnosti so *shape* parametrom $\kappa = 1$ zodpovedá exponenciálnemu rozdeleniu pravdepodobnosti s *rate* parametrom $1/\theta$. Skratka `weibull` reprezentuje v softvéri R Weibullovo rozdelenie pravdepodobnosti a súvisiace funkcie sú analogicky `pweibull`, `dweibull`, `qweibull` a `rweibull`.

2.3 Diskrétne rozdelenia pravdepodobnosti

V tejto podkapitole definujeme len tie diskrétne rozdelenia pravdepodobnosti, s ktorými sa stretneme pri tvorbe aktuárskych modelov v ďalších kapitolách tejto práce. Opäť uvádzame aj niektoré funkcie implementované v softvéri R [9], ktoré súvisia s danými rozdeleniami.

Bernoulliho rozdelenie pravdepodobnosti

Náhodná premenná N sa riadi Bernoulliho rozdelením pravdepodobnosti s parametrom $p \in (0, 1)$, ak má pravdepodobnostnú funkciu [12]:

$$\text{P}(N = n) = p^n (1 - p)^{1-n}, \quad \text{pre } n = 0, 1.$$

Bernoulliho rozdelenie pravdepodobnosti sa nazýva aj alternatívne rozdelenie pravdepodobnosti a označujeme ho symbolom $\text{Alt}(p)$, kde parameter p sa nazýva pravdepodobnostný parameter. Bernoulliho rozdelenie pravdepodobnosti je špeciálnym prípadom binomického rozdelenia pravdepodobnosti. Preto aj v prostredí štatistického softvéru R Bernoulliho (alternatívnemu) rozdeleniu pravdepodobnosti zodpovedá skratka `binom` rovnako, ako aj pre binomické rozdelenie pravdepodobnosti. Funkcie `pbinom`, `dbinom`, `qbinom` a `rbinom` tak vyžadujú nastaviť argument `size` $= 1$, aby sa jednalo o Bernoulliho (alternatívne) rozdelenie pravdepodobnosti.

Poissonovo rozdelenie pravdepodobnosti

Náhodná premenná N sa riadi Poissonovým rozdelením pravdepodobnosti s parametrom $\lambda \in \mathbb{R}^+$, ak má pravdepodobnostnú funkciu [1], [12]:

$$\text{P}(N = n) = \lambda^n \frac{\text{e}^{-\lambda}}{n!}, \quad \text{pre } n \in \mathbb{N}_0.$$

Poissonovo rozdelenie pravdepodobnosti označujeme symbolom Poisson(λ), kde parameter λ sa nazýva *rate* parameter. V prostredí štatistického softvéru R sa Poissonovo rozdelenie pravdepodobnosti označuje ako `pois` a opäť sú v softvéri implementované aj funkcie súvisiace s týmto rozdelením: `ppois`, `dpois`, `qpois` a `rpois`.

V súvislosti s Poissonovým rozdelením pravdepodobnosti definujme aj zložené Poissonovo rozdelenie. Ak náhodná premenná S vznikne ako súčet nezávislých, rovnako rozdelených náhodných premenných $X_1, X_2, X_3, \ldots$, pričom počet týchto sčítancov je náhodná premenná N, ktorá sa riadi Poissonovým rozdelením pravdepodobnosti s parametrom $\lambda \in \mathbb{R}^+$, tak hovoríme, že náhodná premenná S má zložené Poissonovo rozdelenie. Zložené Poissonovo rozdelenie pravdepodobnosti pre náhodnú premennú S budeme označovať zápisom $S \sim \text{CoPoisson}(\lambda, F_X)$, kde F_X zodpovedá distribučnej funkcii rozdelenia pravdepodobnosti náhodných premenných $X_1, X_2, X_3, \ldots$ Zaveďme na záver ešte označenie pre rozdelenie pravdepodobnosti náhodnej premennej Z, ktorá vznikne súčinom náhodných premenných X a N. Predpokladajme, že náhodná premenná X sa riadi napríklad exponenciálnym rozdelením pravdepodobnosti s parametrom $\lambda \in \mathbb{R}^+$, zatiaľ čo náhodná premenná N sa riadi napríklad Bernoulliho (alternatívnym) rozdelením pravdepodobnosti s parametrom $p \in (0, 1)$. Potom pre takéto zmiešané rozdelenie pravdepodobnosti náhodnej premennej Z budeme používať označenie $Z \sim \text{Alt/Exp}(p, \lambda)$.

2.4 Testy dobrej zhody

Medzi známe štatistické metódy využívané pri kalibrácii i validácii aktuárskych modelov patria testy dobrej zhody, ku ktorým radíme napríklad Kolmogorovov-Smirnovov test, Cramérov-von Misesov test a Andersonov-Darlingov test. Základ testov dobrej zhody zvyčajne spočíva v porovnávaní empirickej distribučnej funkcie a teoretickej distribučnej funkcie, pričom pri všetkých zo spomínaných testov sa predpokladá, že náhodné premenné, ktorých empirickú a teoretickú distribučnú funkciu porovnávame, sú nezávislé, rovnako rozdelené a pochádzajú z nejakého spojitého rozdelenia pravdepodobnosti. Nech teda F zodpovedá teoretickej kumulatívnej distribučnej funkcii a nech F_n označuje empirickú kumulatívnu distribučnú funkciu skonštruovanú na základe dostupných dát, pričom n predstavuje veľkosť náhodného výberu. Testy dobrej zhody

potom môžeme formulovať aj pomocou nasledujúcej hypotézy:

$$\mathrm{H}_0 : F_n = F \quad \text{vs.} \quad \mathrm{H}_1 : F_n \neq F.$$

Všetky tri vybrané testy dobrej zhody vychádzajú z rovnakého princípu, ktorým je to, že za testovú štatistiku sa považuje určitý typ vzdialenosti medzi pozorovanými distribučnými funkciami.

Kolmogorovov-Smirnovov test

Nech platia všetky označenia uvedené vyššie. Jednovýberový Kolmogorovov-Smirnovov test, používaný na testovanie nulovej hypotézy H_0, má testovú štatistiku založenú na maximálnej odchýlke medzi funkčnými hodnotami empirickej a teoretickej kumulatívnej distribučnej funkcie, čomu zodpovedá zápis:

$$D = \sup_{x \in \mathcal{D}(F)} |F_n(x) - F(x)|, \tag{2.7}$$

kde $\mathcal{D}(F)$ je definičný obor teoretickej kumulatívnej distribučnej funkcie. Testovú hypotézu H_0 zamietame na hladine významnosti α, ak vzdialenosť medzi empirickou a teoretickou kumulatívnou distribučnou funkciou je „príliš veľká“, čiže vtedy, keď hodnota testovej štatistiky D je väčšia ako tabuľková kritická hodnota Kolmogorovovho-Smirnovovho testu. Kolmogorovov-Smirnovov test je implementovaný aj v štatistickom softvéri R napríklad v balíku `goftest` [4] ako funkcia `ks.test`, ktorej výstupom je hodnota testovej štatistiky D a tiež p-hodnota.

Cramérov-von Misesov test

Opäť predpokladajme, že platia všetky vyššie zavedené označenia. Základom testovej štatistiky jednovýberového Cramérovho-von Misesovho testu je meranie veľkosti plochy, ktorá vzniká medzi empirickou distribučnou funkciou a teoretickou distribučnou funkciou, a teda tvar testovej štatistiky je nasledovný:

$$\omega^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \, \mathrm{d}F(x). \tag{2.8}$$

Testová štatistika Cramérovho-von Misesovho testu sa zvykne uvádzať aj v iných tvaroch, no myšlienka merania veľkosti plochy ostáva nezmenená. Nulovú hypotézu zamietame na hladine významnosti α vtedy, keď hodnota testovej štatistiky Cramérovho-von Misesovho testu je väčšia ako kritická hodnota uvádzaná v tabuľkách pre Cramérov-

von Misesov test. Hodnotu testovej štatistiky ω^2 a p-hodnotu Cramérovho-von Misesovho testu je opäť možné jednoducho získať v prostredí softvéru R pomocou funkcie `cvm.test`, ktorá sa nachádza v balíku `goftest` [4].

Andersonov-Darlingov test

Nakoniec pri nezmenených označeniach uvádzame aj testovú štatistiku Andersonovho-Darlingovho testu, ktorá vznikla modifikáciou testovej štatistiky Cramérovho-von Misesovho testu:

$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} \mathrm{d}F(x). \tag{2.9}$$

Ide tak opäť o meranie veľkosti plochy medzi pozorovanými distribučnými funkciami a navyše Andersonov-Darlingov test kladie vyššiu váhu na chvosty rozdelenia ako Kolmogorovov-Smirnovov test. Testová štatistika A^2 sa opäť zvykne uvádzať aj v iných tvaroch, ktoré však v tejto diplomovej práci neuvádzame. Aj v tomto prípade sa nulová hypotéza zamieta na hladine významnosti α práve vtedy, keď hodnota testovej štatistiky Andersonovho-Darlingovho testu je väčšia ako tabuľková kritická hodnota tohto testu. Pre výpočet hodnoty testovej štatistiky a p-hodnoty Andersonovho-Darlingovho testu je možné využiť funkciu implementovanú v prostredí softvéru R `ad.test`, ktorú je znova možné nájsť v balíku `goftest` [4].

Ako sme už vyššie naznačili, testy dobrej zhody je možné využiť pri validácii modelu. Na následné meranie kvality modelu, respektíve na vzájomné porovnávanie presnosti a vhodnosti odhadnutých modelov, môžeme využiť ukazovatele, ktorými sú Bayesovo informačné kritérium a Akaikeho informačné kritérium. Platí, že za najvhodnejší model sa považuje ten, ktorý dosahuje najnižšiu hodnotu Bayesovho či Akaikeho informačného kritéria zo všetkých porovnávaných modelov.

Bayesovo informačné kritérium

Bayesovo informačné kritérium, ktoré je často označované skrátene aj ako BIC, je definované nasledovne:

$$\mathrm{BIC} = k\ln(n) - 2\ln(\mathcal{L}), \tag{2.10}$$

kde k reprezentuje počet odhadovaných parametrov modelu, n zodpovedá veľkosti náhodného výberu a $\mathcal{L}$ značí funkciu vierohodnosti.

Akaikeho informačné kritérium

Akaikeho informačné kritérium, skrátene AIC, je špeciálnym prípadom Bayesovho informačného kritéria, kedy $\ln(n) = 2$, a tak pre AIC píšeme:

$$\mathrm{AIC} = 2k - 2\ln(\mathcal{L}). \tag{2.11}$$

Hodnoty AIC a BIC je možné získať v štatistickom softvéri R ako výstup z rôznych funkcií. Jednou z nich je funkcia `gofstat` z balíka `fitdistrplus` [2].

Testy dobrej zhody, respektíve ich testové štatistiky využívame aj pri kalibrácii modelu, konkrétne pri odhadovaní parametrov modelu pomocou metódy maximálnej dobrej zhody. Metóda maximálnej dobrej zhody je založená na numerickej optimalizácii hodnoty vybranej účelovej funkcie, kde za túto funkciu sa zvyčajne volí testová štatistika Kolmogorovovho-Smirnovovho testu, Cramérovho-von Misesovho testu alebo Andersonovho-Darlingovho testu [12]. Parametre zvoleného rozdelenia sa následne odhadujú minimalizáciou tejto účelovej funkcie, čiže cieľom je nájsť také parametre, aby daný typ vzdialenosti medzi empirickou a teoretickou distribučnou funkciou bol minimálny. Dodáme, že poznáme niekoľko ďalších metód, ktoré sú používané pri odhadovaní parametrov daného rozdelenia pravdepodobnosti. Medzi ne patrí napríklad metóda maximálnej vierohodnosti, momentová metóda, kvantilová metóda a iné ďalšie metódy, ktorými sa však v tejto práci nezaoberáme.

3 Okruhy matematického modelovania v oblasti aktuárstva

Aktuárske modely sú užitočné vo viacerých oblastiach poisťovníctva, a preto je dobré vedieť takýto model skonštruovať. V tejto kapitole si tak predstavíme postupnosť krokov, ktorá sa zvyčajne používa pri tvorbe aktuárskych modelov. Následne uvedieme niektoré okruhy matematického modelovania v oblasti aktuárstva, ktorým sa budeme konkrétnejšie venovať v jednotlivých podkapitolách tejto kapitoly.

Postupnosť krokov tvorby aktuárskeho modelu

Pri tvorbe aktuárskeho modelu je predovšetkým potrebné zamyslieť sa nad tým, čo vlastne chceme modelovať. Druhou otázkou, ktorú si môžeme položiť, je: prečo to chceme modelovať, prípadne čo je cieľom nášho modelovania. Tento rozbor môžeme považovať za akýsi nultý krok aktuárskeho modelovania. Po zodpovedaní týchto základných otázok nasleduje ďalšia otázka, ktorá znie: ako to chceme modelovať? Tým pristupujeme k samotnému aktuárskemu modelovaniu, ktoré pozostáva z niekoľkých zaužívaných krokov [12]:

- krok 1: získanie, prvotné úpravy a analýza dát,
- krok 2: voľba konkrétneho matematického modelu,
- krok 3: kalibrácia modelu – odhadovanie parametrov zvoleného modelu,
- krok 4: validácia modelu – overovanie vhodnosti modelu,
- krok 5: vyhodnotenie výsledkov.

Napokon podľa [12] doplníme, že proces tvorby aktuárskych modelov je z určitej miery empirickou technikou, teda je potrebné mať určité aktuárske skúsenosti, a preto počet i samotný obsah uvedených krokov nemusí byť pevne stanovený. V ďalšej časti tejto kapitoly si predstavíme vybrané tematické okruhy matematického modelovania v oblasti aktuárstva, pričom vždy analyzujeme, čo chceme modelovať a prečo to chceme modelovať. K otázke, ako to chceme modelovať, sa viaže celá postupnosť krokov tvorby aktuárskeho modelu, a tak sa tejto otázke v nasledujúcich podkapitolách nebudeme konkrétnejšie venovať. Ide o pomerne zdĺhavý postup, ktorý si dopodrobna predstavíme vo štvrtej kapitole, konkrétne pre prípad aktuárskeho modelovania celkovej výšky

poistných nárokov v modeli individuálneho rizika, a tiež v piatej kapitole pri aktuárskom modelovaní procesu rizika v teórii ruinovania. Medzi niekoľko ďalších tematických okruhov aktuárskeho modelovania patrí napríklad aktuárske modelovanie v demografii, finančno-matematické modelovanie výnosových kriviek, modelovanie celkovej výšky poistných nárokov v modeli kolektívneho rizika v neživotnom poistení či modelovanie vývojových trojuholníkových schém používaných pri tvorbe rezerv v neživotnom poistení. Ako prvé si priblížime aktuárske modelovanie v demografii, ktorému sa venujeme v podkapitole 3.1, pričom vychádzame z publikácie [14].

3.1 Aktuárske modelovanie v demografii

Demografia je vedecká disciplína, ktorej predmetom záujmu je analýza štruktúry, veľkosti a rozloženia populácie. Zaoberá sa taktiež zmenami v populácii, ktoré sú spôsobené primárne procesmi, akými sú pôrodnosť, úmrtnosť, sťahovanie a sociálna mobilita. Znalosti z demografie je možné aplikovať aj v oblasti životného poistenia. Je známe, že primárnym rizikom pri uzatváraní zmlúv životného poistenia je z hľadiska poisťovne riziko úmrtia poistených osôb, a tak jednou z hlavných úloh poisťovne je vedieť čo najpresnejšie opísať miery úmrtnosti, čím sa dostávame k aktuárskemu modelovaniu v demografii. Zodpovedajme preto na dve základné otázky.

Čo chceme modelovať?

Ako sme práve naznačili, poisťovňa potrebuje mať k dispozícii čo najpresnejší odhad mier úmrtnosti populácie. Odpoveď na otázku je tak zrejmá, chceme modelovať miery úmrtnosti alebo inými slovami úmrtnostné krivky.

Prečo to chceme modelovať?

Úmrtnostné krivky chceme modelovať preto, aby mala poisťovňa k dispozícii presnejšie informácie o tom, ako sa líši pravdepodobnosť úmrtia v jednotlivých rokoch života osoby, prípadne v jednotlivých vekových skupinách, a tiež ako sa mení v priebehu času. Poisťovňa by potom vedela primerane oceniť svoje poistné produkty či presnejšie stanoviť výšku technických rezerv potrebných na krytie budúcich výdavkov. Skrátene tak môžeme povedať, že cieľom modelovania úmrtnostných kriviek je spresniť odhad finančných rizík spojených so životným poistením.

Tomu, ako konkrétne modelovať úmrtnostné krivky, sa nevenujeme, no uvedieme niekoľko známych demografických modelov používaných pri modelovaní úmrtnostných kriviek, z ktorých je možné si v druhom kroku tvorby aktuárskeho modelu niektorý zvoliť. Jedným z najznámejších modelov je Gompertzov model úmrtnosti, ďalšími známymi modelmi sú napríklad Gompertzov-Makehamov model, Weibullov model, Beardov model a iné.

3.2 Aktuárske modelovanie výnosových kriviek

Pod pojmom výnosová krivka rozumieme vzťah medzi úrokovou mierou a dobou do splatnosti dlhopisov. Výnosová krivka je zväčša prezentovaná graficky, kde na zvislej osi znázorňujeme ročné výnosy dlhopisu a na vodorovnej osi doby do splatnosti dlhopisu, a tak výnosová krivka reprezentuje akúsi časovú štruktúru úrokových mier dlhopisov. Modelovanie výnosových kriviek je v oblasti financií veľmi dôležité. Výnosové krivky nám poskytujú pohľad na celkový stav ekonomiky a na ich základe je možné predikovať budúce úrokové miery. Taktiež vieme, že dlhopisy či opcie sú citlivé na zmenu úrokovej miery, a tak je možné ich na základe predikcií lepšie oceniť. Modelovanie výnosových kriviek má svoj význam aj v oblasti poisťovníctva. Po zodpovedaní dvoch základných otázok zistíme aký.

Čo chceme modelovať?

V tomto prípade je odpoveď jasná – chceme modelovať výnosové krivky.

Prečo to chceme modelovať?

Modelovanie výnosových kriviek je v neposlednom rade dôležité aj v kontexte diskontovania poistných záväzkov v prípade poistných produktov životného poistenia. Poisťovne totiž pri diskontovaní svojich poistných záväzkov aktuálne používajú výnosovú krivku namiesto kedysi používanej konštantnej technickej úrokovej miery [12]. Modelovanie výnosových kriviek napomáha presnejšiemu výpočtu súčasnej hodnoty očakávaných budúcich poistných výdavkov poisťovne, ktorá je za predpokladu takzvaného netto princípu ekvivalencie rovná súčasnej hodnote príjmov poisťovne, kde príjem poisťovne predstavuje finančnú čiastku, ktorú má poistenec zaplatiť tesne po uzatvorení zmluvy vybraného životného poistenia (v prípade jednorazovo plateného poistného).

Na záver uvádzame niektoré z modelov, ktoré sa používajú pri modelovaní výnosových kriviek. Sú nimi napríklad Vašíčkov model, Coxov-Ingersollov-Rossov model (CIR model), Nelsonov-Siegelov model či Svenssonov model.

3.3 Aktuárske modelovanie v modeli kolektívneho rizika

Ďalšou významnou oblasťou aktuárskeho modelovania je modelovanie celkovej výšky poistných nárokov, pričom v tejto podkapitole sa venujeme konkrétne modelu kolektívneho rizika. O predpokladoch tohto modelu si povieme nateraz len málo, no viac sa o ňom dočítame v podkapitole 5.1. V modeli kolektívneho rizika, ktorý je používaný najmä v neživotnom poistení, je celková výška poistných nárokov súčtom všetkých nahlásených poistných nárokov, pričom ich počet je vopred neznámy. Výšky nahlásených poistných nárokov predstavujú nezáporné, nezávislé, rovnako rozdelené náhodné premenné, kde výšku prvého nahláseného poistného nároku môžeme označiť X_1, druhého X_2, tretieho X_3, ... Počet nahlásených poistných nárokov je taktiež náhodná premenná, ktorú môžeme označiť ako N. Zdá sa, že v tomto prípade je potrebné modelovať viacero vecí. Zamyslime sa a odpovedzme opäť na základné otázky.

Čo chceme modelovať?

Tentokrát chceme modelovať celkovú výšku poistných nárokov, ktorá je rovná súčtu všetkých nahlásených poistných nárokov. Avšak, počet nahlásených poistných nárokov nie je pevne stanovený. Potrebujeme preto v prvom rade vedieť samostatne modelovať počet nahlásených poistných nárokov a rovnako aj výšky nahlásených poistných nárokov.

Prečo to chceme modelovať?

Jedným z hlavných dôvodov, prečo chceme modelovať celkovú výšku poistných nárokov, je spresniť odhad budúcich očakávaných výdavkov poisťovne, a s tým je späté aj spresnenie odhadu výšky rezerv vyhradených na vyplácanie poistných plnení.

Ani v tomto prípade neuvádzame odpoveď na otázku, ako presne modelovať počet poistných nárokov, výšku poistných nárokov či celkovú výšku poistných nárokov v modeli kolektívneho rizika. Môžeme však povedať toľko, že z aktuárskych skúseností je známe, že počty nahlásených poistných nárokov sa zvyknú riadiť nejakým diskrétnym rozdele-

ním pravdepodobnosti, zatiaľ čo výšky nahlásených poistných nárokov zväčša sledujú spojité rozdelenia pravdepodobnosti. Viac o modelovaní celkovej výšky poistných nárokov v modeli kolektívneho rizika sa môžeme dočítať v publikácii [12]. Nasledujúca kapitola s názvom *Aktuárske modelovanie v modeli individuálneho rizika* je taktiež venovaná modelovaniu celkovej výšky poistných nárokov, no pre iný model, konkrétne model individuálneho rizika, kde uvádzame celý postup tvorby aktuárskeho modelu spolu s uvedením ilustratívneho príkladu.

3.4 Aktuárske modelovanie vývojových trojuholníkov

Vývojové trojuholníkové schémy predstavujú akúsi trojuholníkovú maticu, ktorá obsahuje informáciu o výškach vyplatených poistných plnení v rôznych časoch. Vyplatené poistné plnenia sú rozdelené do riadkov matice zvyčajne podľa roku vzniku poistnej udalosti a do stĺpcov podľa nasledujúcich rokov vývoja, pričom výšky poistných plnení sú zaznamenané kumulatívne. Takéto trojuholníkové schémy sa zvyknú používať pri tvorbe rezerv v neživotnom poistení. Presuňme sa teda k ich modelovaniu a odpovedzme si na nasledujúce otázky.

Čo chceme modelovať?

Z vyššie uvedeného textu je zrejmé, že chceme modelovať vývojové trojuholníkové schémy.

Prečo to chceme modelovať?

Vývojové trojuholníkové schémy sa používajú na odhad budúcich výšok poistných nárokov, ktoré bude musieť poisťovňa vyplatiť. Tento odhad je opäť užitočný pri oceňovaní poistných zmlúv a pri spomínanej tvorbe rezerv potrebných na krytie budúcich poistných nárokov vyplývajúcich zo vzniknutých poistných udalostí.

Okrem uvedených oblastí aktuárskeho modelovania existuje ešte mnoho ďalších. Z toho čo sme uviedli, však môžeme pozorovať, že základnou myšlienkou aktuárskeho modelovania v akejkoľvek vybranej oblasti je pomáhať poisťovniam a príbuzným organizáciám pri posudzovaní a riadení finančného rizika.

4 Aktuárske modelovanie v modeli individuálneho rizika

Aktuárske alebo poistno-matematické modelovanie má bezpochyby svoje miesto v ekonomickom a finančnom svete. Pomocou rôznych aktuárskych modelov vieme totiž dostatočne presne nastaviť výšky technických rezerv, výšky poistného či odhadnúť počty poistných nárokov. Poistenie je spoľahlivým nástrojom pre vyrovnávanie výšok škôd, keďže hlavnou myšlienkou sektoru poistenia je prenos rizika z jednej zmluvnej strany na druhú, čiže z poistníka na poskytovateľa poistenia, ktorý za túto službu inkasuje poistné. Dôležitou úlohou poisťovní pri tvorbe výšky poistného je identifikácia poistných rizík. Poisťovňa tak potrebuje poznať pravdepodobnosti, s akými môže dôjsť k poistnej udalosti. Rovnako nutné je, aby poisťovňa mala dostatočne presnú informáciu o celkovej výške poistných nárokov, ktoré bude musieť vyplatiť. S cieľom modelovať celkovú výšku poistných nárokov je preto pre poisťovňu nevyhnutné, aby vedela odhadovať výšky poistných nárokov vyplývajúcich z jednotlivých poistných zmlúv a tiež počty poistných nárokov. Pri procese tvorby aktuárskych modelov sa stretávame s rôznymi oblasťami matematického modelovania. Jednu z nich predstavuje matematické modelovanie celkovej výšky poistných nárokov v modeli individuálneho rizika, kde modelu individuálneho rizika je venovaná podkapitola 4.1, pričom vychádzame z publikácii [3], [7], [10] a [13]. V podkapitole 4.2 sa zameriame na postupnosť krokov používaných pri tvorbe aktuárskych modelov, kde postupnosť technických krokov demonštrujeme na riešení ilustratívneho príkladu.

4.1 Model individuálneho rizika

Pri tvorbe pravdepodobnostného modelu na popis celkovej výšky poistných nárokov vyplývajúcich z portfólia poistných zmlúv vychádzame z modelu individuálneho rizika. Ide o štandardný model používaný v poisťovníctve, či už v životnom, neživotnom alebo zdravotnom poistení, pričom my sa venujeme primárne druhému zo spomínaných oblastí poistenia. Model individuálneho rizika je postavený na niekoľkých predpokladoch, ktoré si postupne predstavíme. Majme poistný kmeň tvorený z $n \in \mathbb{N}$ nezávislých poistných zmlúv a predpokladajme, že počet poistných zmlúv ostáva počas

sledovaného obdobia nezmenený. Tieto poistné zmluvy nemusia byť podobné, a tak máme heterogénne poistné portfólio. Na každej poistnej zmluve môže dôjsť počas sledovaného obdobia k najviac jednej poistnej udalosti. Predpokladajme tiež, že výška poistného nároku vyplývajúceho z poistnej zmluvy je rovná výške poistného plnenia. Výšku poistného nároku pri i-tej poistnej zmluve reprezentuje náhodná premenná Z_i, kde $i = 1, 2, 3, \ldots, n$. Je zrejmé, že nie na každej poistnej zmluve musí nastať poistná udalosť, a teda pre niektoré i sa môže stať, že $Z_i = 0$. Celkovú výšku poistných nárokov S^{ind} potom podľa [3], [7] či [13] definujeme nasledovne:

$$S^{\text{ind}} = Z_1 + Z_2 + \cdots + Z_n = \sum_{i=1}^{n} Z_i. \tag{4.1}$$

Celková výška poistných nárokov S^{ind} je súčtom náhodných premenných Z_i, z čoho vyplýva, že aj S^{ind} je náhodná premenná. Na tomto mieste je vhodné uviesť ďalší z kľúčových predpokladov modelu individuálneho rizika, a to ten, že náhodné premenné Z_i opisujúce výšku škody pri i-tej poistnej zmluve nemusia mať rovnaké rozdelenie pravdepodobnosti. Môžu teda nastať dve situácie:

- náhodné premenné Z_i pre $i = 1, 2, 3, \ldots, n$ sa riadia podľa rovnakého zákona rozdelenia pravdepodobnosti, no majú iné hodnoty parametrov,
- náhodné premenné Z_i pre $i = 1, 2, 3, \ldots, n$ sa riadia podľa rôznych rozdelení pravdepodobnosti.

Ako sme už spomínali vyššie, model individuálneho rizika je používaný v životnom, neživotnom a aj v zdravotnom poistení. Najprv uvedieme krátky príklad použitia tohto modelu v oblasti životného poistenia. V životnom poistení sa model používa napríklad pri dočasnom poistení na dožitie alebo pri dočasnom poistení na úmrtie, ak poisťovňa vypláca konštantné poistné plnenie vo výške $y \in \mathbb{R}^+$. Pri dočasnom poistení na dožitie sa poistenej osobe vyplatí poistná suma vo výške y v prípade, že prežije celé poistné obdobie. Ak počas tejto doby zomrie, poistenec alebo oprávnené osoby strácajú nárok na vyplatenie poistnej sumy. V prípade, že neuvažujeme časovú hodnotu peňazí, výšku poistného plnenia (poistnej sumy) pri i-tej poistnej zmluve potom definujeme vzťahom:

$$Z_i = N_i \times y, \tag{4.2}$$

kde N_i je náhodná premenná, ktorá reprezentuje počet poistných nárokov vyplývajúcich z i-tej poistnej zmluvy. V životnom poistení môže takmer vždy dôjsť nanajvýš k jednej poistnej udalosti, čo zodpovedá aj vyššie uvedenému predpokladu, a preto náhodná premenná N_i s určitou pravdepodobnosťou nadobúda hodnoty 0 alebo 1. Ak $N_i = 0$, poistná udalosť nenastala (v našom príklade dočasného poistenia na dožitie poistený zomrel počas poistnej doby a zanikol mu nárok na vyplatenie poistnej sumy) a naopak, ak $N_i = 1$, tak došlo k poistnej udalosti (poistenec prežil celú dobu poistenia a na jej konci mu bude vyplatená poistná suma). Náhodná premenná N_i sa teda riadi Bernoulliho (alternatívnym) rozdelením pravdepodobnosti [10], [13]:

$$N_i = \begin{cases} 0, & \text{s pravdepodobnosťou } 1 - p_i, \\ 1, & \text{s pravdepodobnosťou } p_i, \end{cases} \tag{4.3}$$

kde parameter $p_i \in (0, 1)$ určuje pravdepodobnosť, s akou nastáva poistná udalosť na i-tej poistnej zmluve, pre $i = 1, 2, 3, \ldots, n$. Parameter p_i nazývame aj pravdepodobnostný parameter. V našom príklade dočasného poistenia na dožitie by výška poistnej sumy pri i-tej poistnej zmluve nadobúdala hodnoty:

$$Z_i = \begin{cases} 0, & \text{s pravdepodobnosťou } 1 - p_i, \\ y, & \text{s pravdepodobnosťou } p_i. \end{cases} \tag{4.4}$$

Zložitejšia situácia nastáva, ak výšky poistných nárokov nadobúdajú akékoľvek hodnoty. V tomto prípade je podľa [10] či [13] vhodné náhodnú premennú Z_i modelovať ako súčin dvoch náhodných premenných N_i a Y_i:

$$Z_i = N_i \times Y_i. \tag{4.5}$$

Náhodná premenná N_i opäť reprezentuje počet poistných nárokov vyplývajúcich z i-tej poistnej zmluvy. Vďaka predpokladu, že na každej poistnej zmluve môže počas sledovaného obdobia dôjsť najviac k jednej poistnej udalosti, sa aj teraz náhodná premenná N_i riadi Bernoulliho (alternatívnym) rozdelením pravdepodobnosti, čiže platí vzťah (4.3) a môžeme písať $N_i \sim \text{Alt}(p_i)$. Náhodná premenná Y_i predstavuje výšku nenulového poistného nároku plynúceho z i-tej poistnej zmluvy. K nenulovému poistnému nároku môže dôjsť len za predpokladu, že poistná udalosť nastala, čiže vtedy, keď $N_i = 1$. Náhodná premenná Y_i je teda podmienená náhodná premenná, ktorú značíme aj zápisom $(Y_i|N_i = 1)$. Strednú hodnotu náhodnej premennej Z_i vieme vyjadriť tak, ako aj

v publikáciach [3] a [6], zo vzťahu, ktorého základom je obdoba vety o úplnej strednej hodnote a je tiež známy pod názvom *law of total expectation* alebo *tower rule*:

$$\mathrm{E}(Z_i) = \mathrm{E}(\mathrm{E}(Z_i|N_i)). \tag{4.6}$$

Pomocou vzťahu (4.6) vieme vyjadriť aj disperziu náhodnej premennej Z_i:

$$\mathrm{D}(Z_i) = \mathrm{E}(\mathrm{D}(Z_i|N_i)) + \mathrm{D}(\mathrm{E}(Z_i|N_i)). \tag{4.7}$$

Našim cieľom je však čo najpresnejšie modelovať celkovú výšku poistných nárokov S^{ind}. Chceme odvodiť charakteristiky a pravdepodobnostné rozdelenie tejto náhodnej premennej. Keďže celková výška poistných nárokov je súčtom výšok poistných nárokov vyplývajúcich z jednotlivých zmlúv, pre strednú hodnotu celkovej výšky poistných nárokov potom platí:

$$\mathrm{E}\left(S^{\mathrm{ind}}\right) = \mathrm{E}(Z_1 + Z_2 + \cdots + Z_n) = \mathrm{E}\left(\sum_{n=1}^{n} Z_i\right). \tag{4.8}$$

V modeli individuálneho rizika predpokladáme, že poistné zmluvy sú navzájom nezávislé, a teda aj výšky poistných nárokov vyplývajúce z jednotlivých zmlúv sú vzájomne nezávislé. Vo všeobecnosti platí, že stredná hodnota súčtu nezávislých náhodných premenných sa rovná súčtu ich stredných hodnôt a očakávaná celková výška poistných nárokov je potom:

$$\mathrm{E}\left(S^{\mathrm{ind}}\right) = \mathrm{E}\left(\sum_{n=1}^{n} Z_i\right) = \sum_{n=1}^{n} \mathrm{E}(Z_i) = \mathrm{E}(Z_1) + \mathrm{E}(Z_2) + \cdots + \mathrm{E}(Z_n). \tag{4.9}$$

Rovnako aj pri odvodení vzťahu pre disperziu celkovej výšky poistných nárokov využijeme predpoklad o nezávislosti náhodných premenných Z_i. Z teórie pravdepodobnosti vieme, že kovariancia dvoch nezávislých náhodných premenných je nulová, a teda platí, že disperzia súčtu nezávislých náhodných premenných je súčet ich disperzií. Disperziu celkovej výšky poistných nárokov počítame:

$$\mathrm{D}\left(S^{\mathrm{ind}}\right) = \mathrm{D}\left(\sum_{n=1}^{n} Z_i\right) = \sum_{n=1}^{n} \mathrm{D}(Z_i) = \mathrm{D}(Z_1) + \mathrm{D}(Z_2) + \cdots + \mathrm{D}(Z_n). \tag{4.10}$$

Keďže vo vzťahoch (4.9) a (4.10) vystupujú stredné hodnoty resp. disperzie výšok poistných nárokov vyplývajúcich z jednotlivých poistných zmlúv, je vhodné ich pomocou

vzťahov (4.6) resp. (4.7) dopočítať a spätne dosadiť. Začnime strednou hodnotou:

$$\begin{aligned}\mathrm{E}(Z_i) &= \mathrm{E}(\mathrm{E}(Z_i|N_i)) = \\ &= \mathrm{E}(Z_i|N_i=0) \times \mathrm{P}(N_i=0) + \mathrm{E}(Z_i|N_i=1) \times \mathrm{P}(N_i=1) = \\ &= 0 \times (1-p_i) + \mathrm{E}(Z_i|N_i=1) \times p_i = \mathrm{E}(Y_i|N_i=1) \times p_i. \end{aligned} \tag{4.11}$$

Označme $\mathrm{E}(Y_i|N_i=1) = \mu_i$ a dosaďme do vzťahu (4.11):

$$\mathrm{E}(Z_i) = \mu_i p_i. \tag{4.12}$$

Podobne dopočítame aj disperziu výšok poistných nárokov prislúchajúcich k i-tej poistnej zmluve:

$$\begin{aligned}\mathrm{D}(Z_i) &= \mathrm{E}(\mathrm{D}(Z_i|N_i)) + \mathrm{D}(\mathrm{E}(Z_i|N_i)) = \\ &= \mathrm{D}(Z_i|N_i=0) \times \mathrm{P}(N_i=0) + \mathrm{D}(Z_i|N_i=1) \times \mathrm{P}(N_i=1) + \mathrm{D}(\mathrm{E}(Z_i|N_i)) = \\ &= 0 \times (1-p_i) + \mathrm{D}(Z_i|N_i=1) \times p_i + \mathrm{D}(\mathrm{E}(Z_i|N_i)) = \\ &= \mathrm{D}(Y_i|N_i=1) \times p_i + \mathrm{D}(\mathrm{E}(Z_i|N_i)). \end{aligned} \tag{4.13}$$

Označme $\mathrm{D}(Y_i|N_i=1) = \sigma_i^2$ a dosaďme do vzťahu (4.13):

$$\mathrm{D}(Z_i) = \sigma_i^2 p_i + \mathrm{D}(\mathrm{E}(Z_i|N_i)). \tag{4.14}$$

Vo vzťahu (4.14) ostal ešte neznámy výraz $\mathrm{D}(\mathrm{E}(Z_i|N_i))$. Zamerajme sa najskôr na člen vo vnútri, a to $\mathrm{E}(Z_i|N_i)$. Z vyššie uvedeného predpokladu vieme, že náhodná premenná N_i nadobúda len hodnoty 0 a 1. Takže sa dostávame k dvom prípadom: $\mathrm{E}(Z_i|N_i=0) = 0$ alebo $\mathrm{E}(Z_i|N_i=1) = \mathrm{E}(Y_i|N_i=1) = \mu_i$. Výsledok $\mathrm{E}(Z_i|N_i)$ teda závisí od hodnoty náhodnej premennej N_i, čo môžeme zapísať nasledovne:

$$\mathrm{E}(Z_i|N_i) = \mu_i N_i. \tag{4.15}$$

Pri výpočte disperzie náhodnej premennej Z_i použijeme rovnosť (4.15) a následne využijeme znalosť, že náhodná premenná $N_i \sim \mathrm{Alt}(p_i)$ a jej disperzia $\mathrm{D}(N_i)$ je rovná $p_i(1-p_i)$:

$$\begin{aligned}\mathrm{D}(Z_i) &= \sigma_i^2 p_i + \mathrm{D}(\mu_i N_i) = \sigma_i^2 p_i + \mu_i^2 \times \mathrm{D}(N_i) = \\ &= \sigma_i^2 p_i + \mu_i^2 p_i (1-p_i). \end{aligned} \tag{4.16}$$

Zo vzťahov (4.12) a (4.16) už poznáme, čomu je rovná stredná hodnota resp. disperzia výšky poistných nárokov pre i-tú poistnú zmluvu, a tak tieto vzťahy môžeme dosadiť do rovníc (4.9) a (4.10). Získame tým vzťah pre výpočet strednej hodnoty celkovej výšky poistných nárokov:

$$\mathrm{E}\left(S^{\mathrm{ind}}\right) = \sum_{n=1}^{n} \mathrm{E}(Z_i) = \sum_{n=1}^{n} \mu_i p_i \tag{4.17}$$

a rovnako aj vzťah pre výpočet disperzie celkovej výšky poistných nárokov:

$$\mathrm{D}\left(S^{\mathrm{ind}}\right) = \sum_{n=1}^{n} \mathrm{D}(Z_i) = \sum_{n=1}^{n} \left(\sigma_i^2 p_i + \mu_i^2 p_i (1 - p_i)\right). \tag{4.18}$$

V nasledujúcej podkapitole si na riešení ilustratívneho príkladu ukážeme kroky tvorby aktuárskeho modelu, pričom sa pokúsime získať distribučnú funkciu celkovej výšky poistných nárokov, pomocou ktorej sme schopní určiť napríklad výšku kapitálu potrebného na krytie týchto nárokov, pri ktorej s 99,5-percentnou istotou celková výška poistných nárokov neprekročí túto sumu.

4.2 Tvorba aktuárskeho modelu v modeli individuálneho rizika

Pre každú poisťovňu resp. zaisťovňu je okrem iného nesmierne dôležité spoľahlivo a presne odhadovať celkovú výšku poistných nárokov vyplývajúcu z uzavretých zmlúv. Pre efektívny a správny chod poisťovne si nemôžeme dovoliť vyhradiť na vyplácanie poistných nárokov príliš veľkú rezervu, no na druhej strane nemôžeme výšku vyhradenej rezervy ani podhodnotiť, pretože môže dôjsť k situácii, v ktorej poisťovňa nebude mať z čoho kryť poistné nároky. Našim cieľom je preto zostrojiť aktuársky model, na základe ktorého čo najpresnejšie odhadneme celkovú výšku poistných nárokov. Jednotlivé kroky postupu tvorby aktuárskeho modelu v neživotnom poistení si predstavíme na riešení nasledujúceho ilustratívneho príkladu, pričom vychádzame z poznámok [12].

Príklad 4.1. (*aktuárske modelovanie celkovej výšky poistných nárokov v modeli individuálneho rizika v neživotnom poistení*)

Uvažujme poisťovňu, ktorá ponúka svojim klientom poistenie elektronických zariadení, čím kryje riziká súvisiace so zničením, krádežou či stratou elektroniky. Poisťovňa uzavrela $n = 1000$ heterogénnych poistných zmlúv o poistení elektronických zariadení, ktoré

následne zatriedila do štyroch homogénnejších skupín, v rámci ktorých sú pravdepodobnosti nahlásenia poistného nároku rovnaké. V prvej skupine sa nachádza $n_1 = 190$ poistných zmlúv, v druhej skupine ich je $n_2 = 280$, v tretej $n_3 = 290$ a do poslednej skupiny je zaradených $n_4 = 240$ poistných zmlúv. Predpokladáme, že pri každej poistnej zmluve môže dôjsť maximálne k jednej poistnej udalosti – elektronické zariadenie môže byť zničené, odcudzené či stratené nanajvýš jedenkrát (v prípade krádeže a straty neuvažujeme možnosť, že by sa elektronické zariadenie našlo a mohlo by byť tak znovu odcudzené či stratené). Ukážku o výškach poistných nárokov vyplývajúcich z jednotlivých poistných zmlúv môžeme vidieť v Tabuľke 1. Výšky poistných nárokov sú nahlásené a zaznamenané v eurách. Celý dátový súbor je k dispozícii na adrese: `http://www.iam.fmph.uniba.sk/ospm/data/04_data.txt`. Poisťovňa pracuje s modelom individuálneho rizika a jej cieľom je čo najpresnejšie odhadovať celkovú výšku poistných nárokov.

Tabuľka 1: Ukážka dátového súboru o výškach poistných nárokov vyplývajúcich z jednotlivých poistných zmlúv zaradených do štyroch rôznych skupín.

poradové číslo poistnej zmluvy	1. skupina	2. skupina	3. skupina	4. skupina
1	0,00	0,00	0,00	0,00
2	0,00	0,00	0,00	516,97
3	0,00	568,80	0,00	0,00
4	0,00	458,90	0,00	0,00
5	1054,90	212,67	401,02	576,29
⋮	⋮	⋮	⋮	⋮
288	NA	NA	0,00	NA
289	NA	NA	608,75	NA
290	NA	NA	866,06	NA

(*zdroj*: *vlastné spracovanie*)

Riešenie.

Predtým, než začneme s tvorbou konkrétneho aktuárskeho modelu, je dobré zamyslieť sa nad tým, čo chceme modelovať a prečo to chceme modelovať. V našom prípade je odpoveď nasledovná: chceme modelovať celkovú výšku poistných nárokov vyplývajúcich z uzavretých zmlúv, a to z dôvodu, aby poisťovňa vedela čo najpresnejšie nastaviť výšku kapitálu potrebného na krytie týchto nárokov. Celkovú výšku poistných nárokov vieme odhadnúť na základe nahlásených poistných nárokov z minulých období. Tie máme k dispozícii, a teda tvorbu aktuárskeho modelu začneme analýzou dát.

Krok 1: získanie, načítanie a analýza dát

V Tabuľke 1 pozorujeme nahlásené výšky poistných nárokov za minulé obdobie, no v dolnej časti tabuľky sa nachádzajú chýbajúce hodnoty. Nie je to však taký problém, ako by sa mohlo zdať, dôvod chýbajúcich hodnôt je ten, že v každej zo štyroch skupín je uzatvorený rozdielny počet zmlúv. Čiže v skutočnosti nejde o chýbajúce hodnoty v zmysle nezaevidovaných poistných nárokov, ale len o menší počet uzavretých poistných zmlúv v danej skupine. Vidíme tiež viaceré nulové výšky poistných nárokov, čo je typické pre model individuálneho rizika, keďže v tomto modeli sa pozeráme na každú zmluvu zvlášť. Nulové hodnoty v dátovom súbore ale skresľujú opisné charakteristiky: strednú hodnotu, koeficient šikmosti, koeficient špicatosti či iné. Zdá sa teda, že by bolo vhodné tieto hodnoty oddeliť a venovať sa nenulovým výškam poistných nárokov. Prvky z Tabuľky 1 môžeme zapísať podľa rovnice (4.5) v tvare:

$$Z_{k,j} = N_{k,j} \times Y_{k,j}, \tag{4.19}$$

kde index $k \in \mathbb{N}$ určuje do ktorej skupiny patrí zmluva, v našom prípade $k = 1, 2, 3, 4$. Podobne index $j \in \mathbb{N}$ zodpovedá poradovému číslu poistnej zmluvy v danej skupine ($j = 1, 2, \ldots, n_k$). Zároveň môžeme označiť:

$$Z_k = Z_{k,1} + Z_{k,2} + \cdots + Z_{k,n_k}.$$

Jednotlivé hodnoty z Tabuľky 1 patriace napríklad do 4. skupiny by sme tak mohli zapísať a interpretovať nasledovne:

- $Z_{4,1}$ = výška poistného nároku vyplývajúca z poistnej zmluvy s poradovým číslom 1 zaradená do štvrtej skupiny; výška tohto poistného nároku je nulová ($Z_{4,1} = 0$), čo značí, že nenastala žiadna poistná udalosť ($N_{4,1} = 0$),

- $Z_{4,2}$ = výška poistného nároku vyplývajúca z poistnej zmluvy s poradovým číslom 2 zaradená do štvrtej skupiny; výška tohto poistného nároku je nenulová ($Z_{4,2} = 516{,}97$), čo značí, že došlo k poistnej udalosti, a teda platí, že $N_{4,2} = 1$ a $Z_{4,2} = Y_{4,2}$,
- ...

Nenulové výšky poistných nárokov tak zodpovedajú náhodným premenným $Y_{k,j}$. S cieľom modelovať celkovú výšku poistných nárokov S^{ind}, musíme teda najskôr čo najlepšie modelovať náhodné premenné $Z_{k,j}$, a to pomocou premenných $N_{k,j}$ a $Y_{k,j}$.

Modelovanie počtu poistných nárokov vyplývajúcich z j-tej zmluvy $N_{k,j}$

Pravdepodobnostné rozdelenie počtu poistných nárokov v daných skupinách už poznáme. Je to Bernoulliho (alternatívne) rozdelenie pravdepodobnosti s jedným parametrom $p_k \in (0,1)$, ktorý nazývame aj pravdepodobnostný parameter. Parameter p_k reprezentuje pravdepodobnosť, s akou dôjde k poistnej udalosti v k-tej pozorovanej skupine, prípadne ho môžeme interpretovať aj inými slovami ako pravdepodobnosť nahlásenia poistného nároku v k-tej pozorovanej skupine. Zamerajme sa na odhad hodnoty parametra náhodnej premennej $N_{k,j}$, ktorá v tomto značení reprezentuje počet poistných nárokov vzťahujúcich sa na j-tú zmluvu v k-tej pozorovanej skupine. Hodnotu parametra p_k je možné odhadnúť z daných dát prostredníctvom vzťahu:

$$p_k = \frac{\#(Y_{k,j}|N_{k,j} = 1)}{n_k}, \tag{4.20}$$

kde symbol $\#(Y_{k,j}|N_{k,j} = 1)$ označuje počet nenulových výšok poistných nárokov v k-tej skupine a n_k reprezentuje počet uzavretých poistných zmlúv v k-tej skupine. Začnime odhadom pravdepodobnostného parametra v prípade prvej pozorovanej skupiny. Hodnota odhadnutého pravdepodobnostného parametra p_1 náhodnej premennej $N_{1,j}$, pre $j = 1, 2, 3, \ldots, 190$ je:

$$p_1 = \frac{\#(Y_{1,j}|N_{1,j} = 1)}{n_1} = 0{,}1579.$$

Analogicky by sme odhadli aj pravdepodobnostné parametre p_2, p_3 a p_4. Pre náhodnú premennú $N_{1,j}$ tak platí: $N_{1,j} \sim \text{Alt}(0{,}1579)$. Rozdelenie počtu poistných nárokov aj s jeho odhadnutým parametrom v prípade prvej poistnej skupiny už máme k dispozícii. Môžeme tak prejsť k modelovaniu náhodnej premennej $Y_{k,j}$.

Modelovanie nenulových výšok poistných nárokov $Y_{k,j}$

Rozdelenie pravdepodobnosti nenulových výšok poistných nárokov je narozdiel od rozdelenia pravdepodobnosti počtu poistných nárokov neznáme. Pristúpme preto k ďalšej časti tvorby aktuárskeho modelu.

Krok 2: voľba konkrétneho matematického modelu

Pri modelovaní nenulovej výšky poistných nárokov máme na výber z dvoch prístupov, ktorými sú parametrický a neparametrický prístup. Ako sme už spomenuli vyššie, poznáme pravdepodobnostné rozdelenie náhodnej premennej $N_{k,j}$, čiže v našom prípade je vhodné modelovať aj nenulové výšky poistných nárokov $Y_{k,j}$ parametricky. Hľadáme teda vhodné rozdelenie pravdepodobnosti, ktorým čo najlepšie opíšeme správanie nenulových výšok poistných nárokov. Opäť začneme analýzou nenulových výšok poistných nárokov vyplývajúcich zo zmlúv zaradených do prvej skupiny. V Tabuľke 2 môžeme vidieť základné charakteristiky náhodnej premennej $Y_{1,j}$.

Tabuľka 2: Základné opisné štatistiky nenulových výšok poistných nárokov vyplývajúcich z poistných zmlúv zaradených do 1. skupiny.

minimálna hodnota	maximálna hodnota	stredná hodnota	smerodajná odchýlka	koeficient šikmosti	koeficient excesu
498,75	1354,65	905,06	185,73	0,32	0,11

(*zdroj*: *vlastné spracovanie*)

Zistili sme, že stredná výška jedného poistného nároku v prvej skupine je 905,06 eura. K dispozícii máme aj informáciu o maximálnej výške poistného nároku, z čoho môžeme usúdiť, že medzi výškami poistných nárokov sa nevyskytuje žiadna extrémne vysoká hodnota. Rozdelenie pravdepodobnosti, pomocou ktorého budeme modelovať nenulové výšky poistných nárokov, by preto malo mať ľahký pravý chvost (chvost zhora ohraničený exponenciálnym rozdelením pravdepodobnosti). Ďalšiu znalosť o vhodnom rozdelení pravdepodobnosti nám poskytuje koeficient šikmosti. Jeho hodnota je kladná, čo naznačuje, že by sme mali použiť pozitívne zošikmené rozdelenie pravdepodobnosti. Taktiež si treba uvedomiť, že výšky poistných nárokov nadobúdajú len kladné hodnoty (prípadne nulové, ak nedôjde k žiadnej poistnej udalosti). Preto je na modelovanie vý-

šok poistných nárokov vhodné použiť spojité rozdelenia pravdepodobnosti definované na množine nezáporných reálnych čísel. Skúsme však aj napriek tomu zvoliť na modelovanie výšok poistných nárokov normálne rozdelenie pravdepodobnosti. Je zrejmé, že výsledok nebude postačujúci, no môžeme ho použiť na porovnanie. V ďalšej časti aktuárskeho modelovania budeme model kalibrovať.

Krok 3: odhadovanie parametrov zvoleného modelu

Na základe predchádzajúceho kroku, ktorým je voľba konkrétneho matematického modelu, pre nenulové výšky poistných nárokov volíme:

- **normálne rozdelenie pravdepodobnosti**.

Predpokladáme teda, že $Y_{1,j} \sim \mathcal{N}(\mu_1, \sigma_1^2)$, kde $\mu_1 \in \mathbb{R}$ a $\sigma_1 \in \mathbb{R}^+$. Parametre normálneho rozdelenia pravdepodobnosti odhadneme v prostredí štatistického softvéru R [9] pomocou funkcie `fitdist` z balíka `fitdistrplus` [2], kde hlavička tejto funkcie má tvar:

```
fitdist(data, distr, method = c("mle", "mme", "qme", "mge", "mse"),
        discrete, ...).
```

Do prvého argumentu s názvom `data` zadáme vektor nahlásených nenulových výšok poistných nárokov v 1. skupine. V argumente `distr` uvádzame zvolené rozdelenie pravdepodobnosti. Keďže sme sa rozhodli požiť normálne rozdelenie do funkcie `fitdist` uvedieme v tomto prípade `distr="norm"`. Pomocou argumentu `method` zvolíme, aká metóda má byť použitá: metóda maximálnej vierohodnosti, momentová metóda, kvantilová metóda, metóda maximálnej dobrej zhody a podobne. V našom prípade odhadovania parametrov volíme metódu maximálnej dobrej zhody založenú na Cramérovej-von Misesovej vzdialenosti. Ak teda zvolíme `method="mge"`, je potrebné uviesť aj argument `gof`, v ktorom určíme, aká štatistika má byť použitá. Argument `discrete` má binárny charakter. Ak je jeho hodnota `FALSE`, predpokladá sa, že dáta pochádzajú zo spojitého rozdelenia pravdepodobnosti. V opačnom prípade sa predpokladá, že dáta sú z diskrétneho rozdelenia. Hlavička funkcie `fitdist` má v prípade odhadu parametrov normálneho rozdelenia pravdepodobnosti pomocou metódy maximálnej dobrej zhody pre nenulové výšky poistných nárokov v prvej skupine tvar:

```
fitdist(Y1, distr="norm", method="mge", discrete=FALSE, gof="CvM").
```

Hodnoty odhadnutých parametrov sú $\mu_1 = 898{,}2836$ a $\sigma_1 = 175{,}3678$. Model založený na normálnom rozdelení ale určite nebude ten najvhodnejší, keďže ako sme spomínali vyššie, rozumné je použiť spojité rozdelenia pravdepodobnosti definované na množine nezáporných reálnych čísel. Taktiež vieme, že normálne rozdelenie je symetrické, čomu však protirečí hodnota koeficientu šikmosti z Tabuľky 2. O vhodnosti resp. nevhodnosti modelu sa presvedčíme v ďalšom kroku aktuárskeho modelovania, v ktorom sa venujeme aj meraniu kvality modelu.

Krok 4: validácia a overovanie vhodnosti modelu

Na overovanie vhodnosti aktuárskeho modelu používame testy dobrej zhody, napríklad Kolmogorovov-Smirnovov test, Cramérov-von Misesov test a Andersonov-Darlingov test. V prostredí štatistického softvéru R je v balíku `fitdistrplus` dostupná aj funkcia s názvom `gofstat`, ktorej výsledkom sú hodnoty spomínaných troch testových štatistík [2]. Taktiež nás zaujímajú p-hodnoty týchto testových štatistík, pretože aj na ich základe vieme v konečnom dôsledku určiť vhodnosť modelu. Tie vieme vypočítať opäť v softvéri R pomocou balíku `goftest`. Konkrétne sú na to určené funkcie `ks.test`, `cvm.test` respektíve `ad.test`. Viac informácii o tomto balíku vieme nájsť v programe R pomocou funkcie `help` [9]. V prípade normálneho rozdelenia vyšli p-hodnoty všetkých troch testových štatistík vysoko nad 5 %, z čoho sa zdá, že model založený na normálnom rozdelení je dostatočne presný a vhodný. Funkcia `gofstat` poskytuje okrem hodnôt testových štatistík aj hodnoty Akaikeho informačného kritéria (AIC), prípadne Bayesovho informačného kritéria (BIC), na základe ktorých vieme porovnať kvalitu viacerých modelov. Dostali sme AIC = 401,7233 a BIC = 404,5256. Momentálne však nemáme k dispozícii žiadne iné modely a nemáme tieto hodnoty s čím porovnať, preto by mohlo byť vhodné skúsiť použiť aj iné rozdelenia pravdepodobnosti, ktoré by rozumne opísali správanie výšok nenulových poistných nárokov.

Model založený na normálnom rozdelení opisuje výšky poistných nárokov dostatočne, no napriek tomu vieme, že tento model nie je najvhodnejší, keďže je definovaný na množine všetkých reálnych čísel. Vráťme sa preto späť ku kroku 2 a skúsme zvoliť:

- **log-normálne rozdelenie pravdepodobnosti**.

Log-normálne rozdelenie je definované len na množine nezáporných reálnych čísel a má dva parametre $\mu_1 \in \mathbb{R}$ a $\sigma_1 \in \mathbb{R}^+$, ktoré aj v tomto prípade odhadneme pomocou funkcie

`fitdist` a metódy maximálnej dobrej zhody s účelovou funkciou Cramérovho-von Misesovho testu. Nenulové výšky poistných nárokov by sa mohli riadiť log-normálnym rozdelením s hodnotami odhadnutých parametrov $\mu_1 = 6{,}7940$ a $\sigma_1 = 0{,}1967$. Opäť je ale potrebné overiť vhodnosť a kvalitu modelu. Na základe Kolmogorovovho-Smirnovovho testu, Cramérovho-von Misesovho testu a Andersonovho-Darlingovho testu vhodnosť modelu nezamietame. P-hodnoty týchto testov sú uvedené v Tabuľke 3. Pozrime sa ešte na hodnotu AIC prípadne BIC. Hodnota AIC je na úrovni 401,8827, čo je o niečo viac než pri normálnom rozdelení. Zdá sa teda, že normálne rozdelenie je vhodnejšie než log-normálne rozdelenie. Problém log-normálneho rozdelenia môže byť ten, že ide o rozdelenie s ťažkým pravým chvostom, zatiaľ čo v prípade našich dát je vhodné použiť rozdelenie pravdepodobnosti s ľahkým chvostom. Zvoľme tentokrát rozdelenie, ktoré je definované na množine nezáporných reálnych čísel a ktoré má ľahký pravý chvost. Takým je napríklad:

- **gama rozdelenie pravdepodobnosti**.

Uvažujme prípad, kedy $Y_{1,j} \sim \text{Gama}(\alpha_1, \beta_1)$, potom parameter tvaru (*shape* parameter) $\alpha_1 \in \mathbb{R}^+$ a *rate* parameter $\beta_1 \in \mathbb{R}^+$ odhadneme pomocou funkcie `fitdist`, a to opäť rovnakou metódou ako v predchádzajúcich prípadoch. Výsledkom sú hodnoty: $\alpha_1 = 26{,}1587$ a $\beta_1 = 0{,}0289$. Čo sa týka vhodnosti modelu, testy dobrej zhody potvrdia jeho vhodnosť, tak ako aj v ostatných prípadoch. Viac nám preto napovie hodnota AIC, ktorá je rovná 401,2886, čo je zatiaľ najmenej v porovnaní z predchádzajúcimi hodnotami AIC. Našli sme rozdelenie pravdepodobnosti, ktoré zatiaľ najpresnejšie opisuje správanie výšok poistných nárokov. Vyskúšajme však zvoliť ešte jedno rozdelenie pravdepodobnosti, ktoré má podobné vlastnosti ako gama rozdelenie, a to:

- **Weibullovo rozdelenie pravdepodobnosti**.

Predpokladajme, že $Y_{1,j} \sim \text{Wei}(\kappa_1, \theta_1)$, pričom *shape* parameter $\kappa_1 \in \mathbb{R}^+$ a *scale* parameter $\theta_1 \in \mathbb{R}^+$. Hodnoty parametrov odhadneme rovnakým spôsobom ako v predošlých prípadoch: $\kappa_1 = 5{,}8607$ a $\theta_1 = 961{,}3036$. Na základe testov dobrej zhody môžeme určiť tento model ako dostatočne vhodný a hodnota AIC = 405,6330, čo je viac než hodnota AIC pre model založený na gama rozdelení pravdepodobnosti. Vhodnosť modelu by sme mohli posúdiť aj na základe grafických techník, akými sú napríklad porovnanie empirickej a teoretickej distribučnej funkcie či hustoty, Q-Q graf, prípadne P-P graf

a podobne. Mohli by sme ešte vyskúšať mnoho ďalších pravdepodobnostných rozdelení s cieľom nájsť to najlepšie, alebo sa uspokojíme s dostatočne presným a kvalitným modelom. Postupne sme sa tak dostali k poslednému kroku tvorby aktuárskych modelov, ktorý v tejto práci uvážime.

Krok 5: vyhodnotenie výsledkov

Ak sa nám model nezdá dostatočne presný a kvalitný, ako to bolo napríklad po modelovaní normálnym rozdelením, tak sa vrátime k druhému kroku, ktorým je voľba konkrétneho matematického modelu a postup zopakujme, až kým model nie je dostatočne vhodný. Krok po kroku tak prichádzame k záveru, že počet poistných nárokov vyplývajúci z j-tej poistnej zmluvy ($j = 1, 2, 3, \ldots, 190$) zaradenej do 1. poistnej skupiny sa riadi Bernoulliho (alternatívnym) rozdelením pravdepodobnosti s parametrom $p_1 = 0{,}1579$, zatiaľ čo pre nenulové výšky poistných nárokov, ktoré vyplývajú z j-tej poistnej zmluvy, pre $j = 1, 2, 3, \ldots, 190$, sa na základe hodnoty AIC javí byť najvhodnejšie gama rozdelenie pravdepodobnosti s parametrami $\alpha_1 = 26{,}1587$ a $\beta_1 = 0{,}0289$. Celkové zhrnutie hodnôt odhadnutých parametrov pre jednotlivé rozdelenia pravdepodobnosti, hodnôt AIC a p-hodnôt testov dobrej zhody pre zmluvy zaradené do 1. skupiny môžeme pozorovať v Tabuľke 3, kde hrubým písmom je zvýraznené rozdelenie, ktoré sa javí byť zo všetkých nami zvolených rozdelení najvhodnejšie na modelovanie nenulovej výšky poistných nárokov v danej skupine.

Tabuľka 3: Zhrnutie odhadnutých parametrov pre nenulové výšky poistných nárokov v 1. skupine, hodnôt AIC a p-hodnôt testov dobrej zhody vybraných rozdelení pravdepodobnosti.

rozdelenie $Y_{1,j}$	odhad parametrov		AIC	K-S test	C-vM test	A-D test
$\mathcal{N}(\mu_1, \sigma_1^2)$	898,2836	$175{,}3678^2$	401,7233	0,8821	0,9243	0,9005
$\mathrm{LN}(\mu_1, \sigma_1^2)$	6,7940	$0{,}1967^2$	401,8827	0,8732	0,9239	0,9238
$\mathbf{Gama(\alpha_1, \beta_1)}$	**26,1587**	**0,0289**	**401,2886**	**0,8765**	**0,9294**	**0,9365**
$\mathrm{Wei}(\kappa_1, \theta_1)$	5,8607	961,3036	405,6330	0,8960	0,8965	0,6802

(*zdroj*: *vlastné spracovanie*)

Rovnakým spôsobom by sme postupovali aj v prípade 2. poistnej skupiny. Pri voľbe konkrétneho matematického modelu opäť zvolíme rovnaké štyri rozdelenia pravdepodobnosti (mohli by sme však zvoliť aj iné), a následne odhadneme parametre počtu poistných nárokov a nenulových výšok poistných nárokov. Pravdepodobnostný parameter p_2 odhadneme pomocou vzťahu (4.19), pričom odhad je nasledovný: $p_2 = 0{,}3250$. Pri konštrukcii a validácii modelu pre nenulové výšky poistných nárokov $Y_{2,j}$, pre $j = 1, 2, 3, \ldots, 280$, sme zistili, že $Y_{2,j}$ sa riadi Weibullovým rozdelením pravdepodobnosti s parametrami $\kappa_2 = 3{,}1943$ a $\theta_2 = 479{,}4815$. V Tabuľke 4 môžeme pozorovať výsledky aj pre ostatné zvolené rozdelenia.

Tabuľka 4: Zhrnutie odhadnutých parametrov pre nenulové výšky poistných nárokov v 2. skupine, hodnôt AIC a p-hodnôt testov dobrej zhody vybraných rozdelení pravdepodobnosti.

rozdelenie $Y_{2,j}$	odhad parametrov		AIC	K-S test	C-vM test	A-D test
$\mathcal{N}(\mu_2, \sigma_2^2)$	427,9021	$152{,}7424^2$	1176,1426	0,9800	0,9740	0,9528
$\mathrm{LN}(\mu_2, \sigma_2^2)$	6,0376	$0{,}3663^2$	1188,0993	0,5570	0,4745	0,1811
$\mathrm{Gama}(\alpha_2, \beta_2)$	7,6905	0,0175	1178,6928	0,8083	0,6804	0,5001
Wei(κ_2, θ_2)	**3,1943**	**479,4815**	**1174,6822**	**0,9749**	**0,9670**	**0,9391**

(*zdroj*: *vlastné spracovanie*)

Aj v prípade tretej a štvrtej poistnej skupiny zvolíme na základe predošlých skúseností rovnaké štyri rozdelenia pravdepodobnosti a prejdeme ku kalibrácii a následnej validácii modelov. Dostávame odhad pravdepodobnostných parametrov $p_3 = 0{,}1621$ a $p_4 = 0{,}2417$. Hodnoty odhadnutých parametrov zvolených rozdelení pravdepodobnosti pre náhodné premenné $Y_{3,j}$, pre $j = 1, 2, 3, \ldots, 290$, a $Y_{4,j}$, pre $j = 1, 2, 3, \ldots, 240$, k ním prislúchajúce hodnoty AIC a p-hodnoty testov dobrej zhody sú spísané v Tabuľke 5 a v Tabuľke 6. Hrubým písmom je v tabuľkách opäť zvýraznené rozdelenie pravdepodobnosti, ktoré považujeme za najvhodnejšie zo všetkých zvolených rozdelení. V 3. skupine sa nenulové výšky poistných nárokov riadia gama rozdelením pravdepodobnosti a v poslednej 4. skupine môžeme na základe hodnoty AIC zvoliť za najvhodnejšie log-normálne rozdelenie pravdepodobnosti.

Tabuľka 5: Zhrnutie odhadnutých parametrov pre nenulové výšky poistných nárokov v 3. skupine, hodnôt AIC a p-hodnôt testov dobrej zhody vybraných rozdelení pravdepodobnosti.

rozdelenie $Y_{3,j}$	odhad parametrov		AIC	K-S test	C-vM test	A-D test
$\mathcal{N}(\mu_3, \sigma_3^2)$	546,4882	$171{,}6197^2$	624,8076	0,8535	0,8762	0,7705
$\mathrm{LN}(\mu_3, \sigma_3^2)$	6,2871	$0{,}3265^2$	622,8396	0,7501	0,8858	0,9132
$\mathbf{Gama(\alpha_3, \beta_3)}$	**9,7395**	**0,0174**	**621,8032**	**0,7939**	**0,9099**	**0,9595**
$\mathrm{Wei}(\kappa_3, \theta_3)$	3,6082	605,5593	625,6689	0,8764	0,8608	0,6795

(*zdroj*: *vlastné spracovanie*)

Tabuľka 6: Zhrnutie odhadnutých parametrov pre nenulové výšky poistných nárokov v 4. skupine, hodnôt AIC a p-hodnôt testov dobrej zhody vybraných rozdelení pravdepodobnosti.

rozdelenie $Y_{4,j}$	odhad parametrov		AIC	K-S test	C-vM test	A-D test
$\mathcal{N}(\mu_4, \sigma_4^2)$	784,8372	$155{,}9955^2$	759,3783	0,9432	0,9598	0,9128
$\mathbf{LN(\mu_4, \sigma_4^2)}$	**6,6590**	$\mathbf{0{,}2027^2}$	**754,7825**	**0,9527**	**0,9648**	**0,9940**
$\mathrm{Gama}(\alpha_4, \beta_4)$	24,7969	0,0313	755,4198	0,9477	0,9737	0,9933
$\mathrm{Wei}(\kappa_4, \theta_4)$	5,7830	839,8933	772,3948	0,9179	0,8672	0,4942

(*zdroj*: *vlastné spracovanie*)

V tejto chvíli už poznáme pravdepodobnostné rozdelenia nenulových výšok poistných nárokov v jednotlivých skupinách a tiež máme k dispozícii odhady parametrov týchto rozdelení. Rovnako poznáme odhady pravdepodobnostného parametra p_k pre $k = 1, 2, 3, 4$. Z rovnice (4.5) resp. (4.19) vieme, že výšky poistných nárokov Z_i modelujeme ako súčin náhodných premenných N_i a Y_i, a preto rozdelenie náhodnej premennej Z_i bude zmiešané z rozdelení týchto dvoch náhodných premenných. Napríklad pre prvú skupinu by sme tak podľa zavedeného označenia mohli písať: $Z_1 \sim \mathrm{Alt/Gama}(p_1, \alpha_1, \beta_1)$. Strednú hodnotu náhodnej premennej Z_k pre jednotlivé

skupiny môžeme jednoducho vypočítať pomocou vzťahu (4.6), ktorý vychádza z obdoby vety o úplnej strednej hodnote, a následne aj jej disperziu, a to zo vzťahu (4.7). Číselné hodnoty tu však neuvádzame, pretože viac ako charakteristiky náhodnej premennej Z_k nás zaujímajú charakteristiky náhodnej premennej S^{ind}, keďže našim cieľom je čo najpresnejšie modelovať celkovú výšku poistných nárokov. Teoretickú strednú hodnotu respektíve disperziu náhodnej premennej S^{ind} môžeme vypočítať zo vzťahov (4.17) a (4.18). Dostávame: $\mathrm{E}\left(S^{\text{ind}}\right) = 138721{,}93$ a $\mathrm{D}\left(S^{\text{ind}}\right) = 77698158$, prípadne aj hodnotu smerodajnej odchýlky $\sigma\left(S^{\text{ind}}\right) = 8814{,}66$. Náhodná premenná S^{ind} sa riadi nejakým netriviálnym rozdelením, ktoré vzniká spojením Bernoulliho (alternatívneho) rozdelenia s gama rozdelením, prípadne Weibullovým rozdelením či log-normálnym rozdelením, a následného súčtu týchto zmiešaných rozdelení. Rozdelenie celkovej výšky poistných nárokov S^{ind} sa preto pokúsime aproximovať pomocou normálneho rozdelenia pravdepodobnosti. To je však možné len za platnosti Ljapunovovej centrálnej limitnej vety, ktorá uvažuje náhodné premenné, ktoré nemusia pochádzať z rovnakého rozdelenia pravdepodobnosti.

Veta 4.1. [10], [11] *(Ljapunovova centrálna limitná veta)*
Nech $Z_1, Z_2, \ldots, Z_n$ je postupnosť nezávislých náhodných premenných, nie nutne rovnako rozdelených so strednými hodnotami $\mathrm{E}(Z_i) = \mu_i$, disperziami $\mathrm{D}(Z_i) = \sigma_i^2$ a nech existuje tretí centrálny moment:

$$k_n^3 = \sum_{i=1}^{n} \mathrm{E}(Z_i - \mu_i)^3.$$

Potom ak platí Ljapunovova podmienka:

$$\lim_{n\to\infty} \frac{k_n}{s_n} = 0,$$

kde

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2,$$

tak platí centrálna limitná veta a náhodnú premennú S^{ind} môžeme aproximovať normálnym rozdelením pravdepodobnosti:

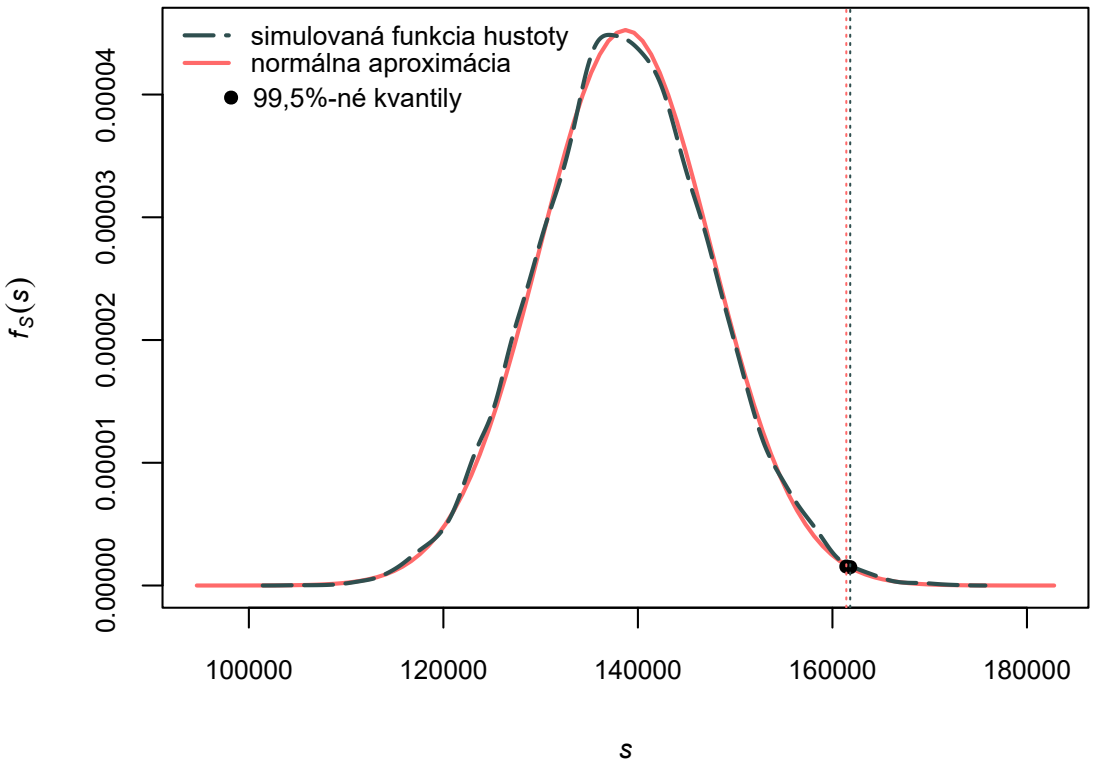
$$\frac{S^{\text{ind}} - \mathrm{E}\left(S^{\text{ind}}\right)}{\sqrt{\mathrm{D}\left(S^{\text{ind}}\right)}} \dot{\sim} \mathcal{N}(0,1).$$

Dôkaz. Dôkaz je možné nájsť v publikácii [11]. □

Aproximáciu celkovej výšky poistných nárokov normálnym rozdelením môžeme jednoducho vykonať v prostredí štatistického softvéru R s využitím platnosti vzťahu:

$$S^{\text{ind}} \dot{\sim} \mathcal{N}\left(\mathrm{E}\left(S^{\text{ind}}\right), \mathrm{D}\left(S^{\text{ind}}\right)\right),$$

kde stredná hodnota a disperzia celkovej výšky poistných nárokov sú už známe. Iný spôsob, ktorým je možné odhadnúť rozdelenie celkovej výšky poistných nárokov, predstavuje použitie rôznych simulačných metód, napríklad metódy Monte Carlo. Simuláciu rozdelenia celkovej výšky poistných nárokov je opäť možné realizovať v prostredí štatistického softvéru R, pričom počet simulačných behov pre jednu Monte Carlo simuláciu sme nastavili na 15000. Výsledok jednej tejto simulácie spolu s výsledkom normálnej aproximácie sme sa rozhodli znázorniť graficky, konkrétne pomocou funkcie hustoty.



Obr. 1: Grafické porovnanie simulovanej funkcie hustoty a normálnej aproximácie hustoty celkovej výšky poistných nárokov (*zdroj: vlastné spracovanie v softvéri R* [9])

Z grafického porovnania hustôt na Obr. 1 môžeme na prvý pohľad usúdiť, že medzi simulovanou funkciou hustoty a normálnou aproximáciou hustoty sú len mierne rozdielnosti, a tak sa zdá, že normálne rozdelenie pravdepodobnosti celkom dobre opisuje dané zložené rozdelenie pravdepodobnosti celkovej výšky poistných nárokov. Na základe hodnôt koeficientov šikmosti a excesu môžeme taktiež vyvodiť isté závery. Je známe, že koeficient šikmosti normálneho rozdelenia je nulový, zatiaľ čo koeficient šikmosti vypočítaný z Monte Carlo simulácie má hodnotu 0,05, čo značí, že rozdelenie je len mierne pozitívne zošikmené. Nulovú hodnotu rovnako nadobúda aj koeficient excesu normálneho rozdelenia. Simulovaný koeficient excesu je rovný hodnote $-0{,}02$, čo naznačuje, že simulované rozdelenie pravdepodobnosti má len o niečo nižšiu mieru špicatosti než normálne rozdelenie. Tieto hodnoty sa líšia od nuly len mierne a nepovažujeme ich žiadnym spôsobom za výrazné. Ďalšie opisné charakteristiky normálnej aproximácie a jednej konkrétnej Monte Carlo simulácie rozdelenia celkovej výšky poistných nárokov sú uvedené v Tabuľke 7.

Tabuľka 7: Opisné charakteristiky a 99,5-percentný kvantil normálnej aproximácie a Monte Carlo simulácie rozdelenia pravdepodobnosti celkovej výšky poistných nárokov.

	Normálna aproximácia	Monte Carlo simulácia
Stredná hodnota	138721,93	138622,90
Smerodajná odchýlka	8814,66	8869,36
Koeficient šikmosti	0,00	0,05
Koeficient excesu	0,00	−0,02
99,5-percentný kvantil	161426,98	161828,03

(*zdroj*: *vlastné spracovanie*)

V Tabuľke 7 uvádzame aj hodnotu 99,5-percentného kvantilu, pretože táto hodnota môže zodpovedať výške rezervy, ktorú by si mala poisťovňa vyhradiť, aby s 99,5-percentnou pravdepodobnosťou pokryla vzniknuté poistné nároky. Tieto hodnoty sú vyznačené aj na Obr. 1, a to čiernymi bodmi a k ním prislúchajúcimi zvislými čiarami, kde červená prerušovaná čiara zodpovedá kvantilu normálnej aproximácie a čierna pre-

rušovaná čiara zase simulovanému kvantilu. Na základe týchto kvantilov by sme mohli konštatovať, že ak by poisťovňa aproximovala celkovú výšku poistných nárokov normálnym rozdelením, pravdepodobne by podhodnotila výšku rezervy, ktorú by bolo potrebné vyhradiť. Nie je však vhodné a postačujúce, vyvodzovať záver na základe porovnania jednej konkrétnej Monte Carlo simulácie s normálnou aproximáciou. V prostredí softvéru R sme preto realizovali ďalšie Monte Carlo simulácie. Celkovo sme sa rozhodli spustiť 50 nezávislých simulácií. Ukazuje sa, že simulované stredné hodnoty a smerodajné odchýlky sa pohybujú okolo hodnôt získaných z normálnej aproximácie. Takisto aj koeficient excesu nadobúda hodnoty okolo nuly. Čo sa týka koeficientu šikmosti, ten je takmer vo všetkých prípadoch kladný, avšak len mierne vyšší od nulovej hodnoty. Zamerali sme sa aj na hodnoty simulovaného 99,5-percentného kvantilu a jeho porovnanie s 99,5-percentným kvantilom normálnej aproximácie. Kvantil normálnej aproximácie je rovný 161426,98 eura a hodnota simulovaného kvantilu bola až v 43 prípadoch simulácií vyššia ako táto hodnota. Zdá sa teda, že normálna aproximácia má skutočne tendenciu podhodnocovať 99,5-percentný kvantil. To by mohlo naznačovať, že pravdepodobnostné rozdelenie celkovej výšky poistných nárokov má o niečo ťažšie chvosty než normálne rozdelenie pravdepodobnosti. Avšak maximálna odchýlka medzi pozorovanými kvantilmi je rovná 1578,78 eura, čo pri hodnote 161426,98 eura nepredstavuje až taký výrazný rozdiel. Môžeme teda povedať, že alternatívne a gama rozdelenie tvoria obstojnú kombináciu a z nich získané netriviálne rozdelenie celkovej výšky poistných nárokov relatívne spoľahlivo konverguje v zmysle Ljapunovovej centrálnej limitnej vety k normálnemu rozdeleniu pravdepodobnosti.

5 Aktuárske modelovanie v teórii ruinovania

Pri tvorbe aktuárskych modelov sa stretávame s ďalšou oblasťou matematického modelovania, a tou je matematické modelovanie procesu rizika v teórii ruinovania. Teória ruinovania sa spája so štúdiom náhodných procesov a zaoberá sa časovým vývojom hodnoty prebytku poisťovne. Základom klasického modelu teórie ruinovania je zovšeobecnenie modelu kolektívneho rizika, ktorý si predstavíme v podkapitole 5.1, pričom vychádzame z publikácii [3], [6] či [12]. V podkapitole 5.2, vychádzajúc z [3], [5], [6], [8] a [13], následne prezentujeme klasický model teórie ruinovania a nakoniec demonštrujeme postup tvorby aktuárskych modelov aj na ilustratívnom príklade.

5.1 Model kolektívneho rizika

Je známych niekoľko modelov, ktoré sa používajú na modelovanie celkovej výšky poistných nárokov. Jedným z nich je model individuálneho rizika, prezentovaný v podkapitole 4.1, ktorý sa zväčša používa v životnom poistení. Iným modelom je model kolektívneho rizika používaný vo väčšine prípadov v neživotnom poistení. V modeli individuálneho rizika sme sa zaoberali každou zmluvou zvlášť, no v modeli kolektívneho rizika sa pozeráme iba na tie zmluvy, pri ktorých došlo k nahláseniu poistného nároku a vyplateniu poistného plnenia. Tieto dva modely sú si v niečom podobné, no majú aj isté odlišnosti. Priblížme si teda model kolektívneho rizika. Uvažujme poistný kmeň tvorený z dostatočne veľkého množstva homogénnych nezávislých poistných zmlúv. Predpokladajme opäť ideálnu situáciu, kde výška poistného nároku je rovná výške poistného plnenia. Nech nezáporná náhodná premenná N predstavuje v prípade modelu kolektívneho rizika počet poistných nárokov nahlásených za jedno časové obdobie, napr. za jeden rok a nech výšku i-teho poistného nároku reprezentuje náhodná premenná X_i, pre $i = 1, 2, \ldots, N$. Celkovú výšku poistných nárokov nahlásených za určité obdobie tak môžeme vyjadriť ako súčet jednotlivých poistných nárokov:

$$S^{\text{kol}} = X_1 + X_2 + \cdots + X_N = \sum_{i=1}^{N} X_i, \tag{5.1}$$

pričom musia byť splnené nasledujúce predpoklady:

- výšky poistných nárokov X_i, pre $i = 1, 2, \ldots$ sú navzájom nezávislé, rovnako rozdelené, nezáporné náhodné premenné,

- výšky poistných nárokov X_i, pre $i = 1, 2, \ldots$ a počty poistných nárokov N sú navzájom nezávislé,
- ak nenastane žiadna poistná udalosť, čiže $N = 0$, tak aj celková výška poistných nárokov $S^{\text{kol}} = 0$.

Náhodná premenná S^{kol} je súčtom nezávislých, rovnako rozdelených náhodných premenných, pričom počet sčítancov je náhodný. V takomto prípade hovoríme, že celková výška poistných nárokov S^{kol} má zložené rozdelenie pravdepodobnosti. Navyše, za predpokladu existencie prvého a druhého počiatočného konečného momentu náhodných premenných $X_i \equiv X$ a $N_i \equiv N$: $\mathrm{E}(X)$, $\mathrm{E}(X^2)$, $\mathrm{E}(N)$ a $\mathrm{E}(N^2)$, vieme vypočítať základné charakteristiky celkovej výšky poistných nárokov, a to konkrétne strednú hodnotu a disperziu náhodnej premennej S^{kol} pomocou Waldových identít [12]:

$$\mathrm{E}\left(S^{\text{kol}}\right) = \mathrm{E}(N)\mathrm{E}(X), \tag{5.2}$$

$$\mathrm{D}\left(S^{\text{kol}}\right) = \mathrm{E}(N)\mathrm{D}(X) + \mathrm{D}(N)\mathrm{E}^2(X). \tag{5.3}$$

Rozdelenie pravdepodobnosti celkovej výšky poistných nárokov S^{kol} je možné opäť aproximovať normálnym rozdelením pravdepodobnosti, prípadne ho môžeme odhadnúť pomocou Monte Carlo simulácií, či využiť iné používané metódy odhadovania rozdelenia pravdepodobnosti.

5.2 Teória ruinovania

V predchádzajúcej podkapitole sme sa oboznámili s modelom kolektívneho rizika, v ktorom pracujeme s poistnými nárokmi nahlásenými za jedno časové obdobie, najčastejšie jeden mesiac alebo jeden rok, a preto tento model môžeme nazývať aj „modelom na krátke obdobie“. Jeho zovšeobecnením sa priblížime k podstate modelov teórie ruinovania, ktorých cieľom je skúmať hodnotu prebytku poisťovne v dlhšom časovom intervale (napríklad niekoľkých rokov). Modely teórie ruinovania sa zaoberajú náhodným vývojom hodnoty prebytku poisťovne v čase, pričom sa sleduje či dôjde k zruinovaniu poisťovne, prípadne s akou pravdepodobnosť sa to stane, a tak opisujú aj stabilitu poisťovne. Ide o náhodný proces, v ktorom začínajúc s určitou počiatočnou nezápornou hodnotou prebytku, sa hodnota prebytku poisťovne lineárne zvyšuje v dôsledku prijatého poistného, no zároveň skokovo klesá v čase, keď nastane poistná udalosť a dôjde

k vyplateniu poistného nároku. Vývoj hodnoty prebytku poisťovne $U(t)$ je možné sledovať v diskrétnom i spojitom čase t, pričom my sa v tejto práci venujeme len modelu so spojitým časom, ktorý sa nazýva aj klasický model teórie ruinovania a je definovaný nasledovne:

$$U(t) = u + ct - S(t), \tag{5.4}$$

kde:

- $U(t)$ – hodnota prebytku poisťovne v čase t,
- $u \equiv U(0)$ – nezáporná počiatočná hodnota prebytku,
- c – nezáporná konštantná výška prijímaného poistného za jednu časovú jednotku,
- $S(t) \equiv S^{\text{kol}}(t)$ – celková výška poistných nárokov nahlásených do času t.

Náhodný proces prebytku poisťovne $\{U(t); t \geq 0\}$ je modelovaný pomocou troch rôznych premenných, no náhodný charakter má len hodnota celkovej výšky poistných nárokov nahlásených do času t. Celkovú výšku poistných nárokov nahlásených za jedno časové obdobie sme schopní modelovať pomocou modelu kolektívneho rizika (5.1). Jeho zovšeobecnením však získame model pre odhad celkovej výšky poistných nárokov nahlásených do času t definovaný predpisom:

$$S(t) = X_1 + X_2 + \cdots + X_{N(t)} = \sum_{i=1}^{N(t)} X_i. \tag{5.5}$$

Ide o náhodný proces celkovej výšky poistných nárokov nahlásených do času t, ktorý modelujeme pomocou postupnosti nezávislých, rovnako rozdelených, nezáporných náhodných premenných X_i, pre $i = 1, 2, \ldots$, ktoré zodpovedajú výške i-teho nahláseného poistného nároku. K i-tej výške poistného nároku X_i navyše definujeme aj čas nahlásenia tohto poistného nároku T_i, pre $i = 1, 2, \ldots$, pričom čas nahlásenia poistného nároku T_i a k nemu priradená výška nahláseného poistného nároku X_i sú navzájom nezávislé. Predpokladajme, že pre postupnosť časových bodov T_i platí, že $T_0 = 0$ a $0 \leq T_1 \leq T_2 \leq \ldots$ Počet nahlásených poistných nárokov do času t je reprezentovaný súborom náhodných premenných $\{N(t); t \geq 0\}$, ktoré nadobúdajú len nezáporné celočíselné hodnoty. Ide teda o náhodný proces počtu nahlásených poistných nárokov do času t, ktorý nazývame aj načítacím procesom a je možné ho definovať zápisom [8]:

$$N(t) = \#\{i \in \mathbb{N} : T_i \leq t\}, \quad \text{pre } t \geq 0. \tag{5.6}$$

Špeciálnym prípadom načítacieho procesu je Poissonov proces, ktorým je aj náhodný proces počtu nahlásených poistných nárokov $\{N(t); t \geq 0\}$.

Definícia 5.1. [6], [13] *(Poissonov proces)*
Načítací proces $\{N(t); t \geq 0\}$ sa nazýva Poissonov proces s *rate* parametrom $\lambda \in \mathbb{R}^+$, ak sú splnené nasledujúce vlastnosti:

1. $N(0) = 0$,
2. prírastky sú nezávislé: pre disjunktné intervaly $(t_i, t_i + h_i)$ sú prírastky procesu $N(t_i + h_i) - N(t_i)$ nezávislé pre všetky hodnoty $h_i > 0$, $i = 1, 2, \ldots$,
3. prírastky sú stacionárne: $N(t+h) - N(t) \sim \text{Poisson}(\lambda h)$ pre všetky hodnoty $t \geq 0$ a $h > 0$.

Navyše platí, že ak je načítací proces počtu poistných nárokov $\{N(t); t \geq 0\}$ Poissonov proces, tak jednotlivé zložky náhodného procesu, reprezentujúce počet poistných nárokov nahlásených do fixného času t, sa riadia Poissonovým rozdelením pravdepodobnosti s parametrom λt. Stredná hodnota a disperzia Poissonovho rozdelenia pravdepodobnosti sú rovnaké, a tak platí, že $\mathrm{E}(N(t)) = \mathrm{D}(N(t)) = \lambda t$, zatiaľ čo kumulatívnu distribučnú funkciu počtu poistných nárokov môžeme vyjadriť nasledovne:

$$\mathrm{P}(N_t \leq n) = \sum_{j=0}^{n} \frac{\mathrm{e}^{-\lambda t}(\lambda t)^j}{j!}, \quad \text{pre } n > 0, \tag{5.7}$$

kde $n \in \mathbb{N}$ značí počet nahlásených poistných nárokov. Vzťah (5.7) môžeme interpretovať aj ako pravdepodobnosť nahlásenia nanajvýš n poistných nárokov do času t. Poissonov proces je možné podľa [5] či [13] definovať aj iným spôsobom, a to pomocou určenia rozdelenia pravdepodobnosti doby uplynutej medzi nahlásením dvoch po sebe nasledujúcich poistných nárokov W_i, pre ktorú platí:

$$W_i = T_i - T_{i-1} \quad \text{resp.} \quad T_i = W_1 + W_2 + \cdots + W_i, \quad \text{pre } i \geq 1. \tag{5.8}$$

Potom, ak sú doby uplynuté medzi nahlásením po sebe nasledujúcich poistných nárokov W_i, pre $i = 1, 2, \ldots$ nezáporné, nezávislé, rovnako rozdelené náhodné premenné riadiace sa exponenciálnym rozdelením pravdepodobnosti s *rate* parametrom $\lambda \in \mathbb{R}^+$, tak náhodný proces počtu poistných nárokov $\{N(t); t \geq 0\}$ sa nazýva Poissonov proces s parametrom λ, kde tento parameter nazývame aj intenzita nahlásenia poistných

nárokov. Pre doby uplynuté medzi nahlásením po sebe nasledujúcich poistných nárokov tiež platí, že sú nezávislé od minulého vývoja procesu. Totiž jednou z vlastností exponenciálneho rozdelenia je, že ide o rozdelenie bez pamäti, a tým sa Poissonov proces zároveň odlišuje od iných načítacích procesov. Okrem toho je známe, že výsledkom súčtu n nezávislých, rovnako rozdelených náhodných premenných, ktoré sa riadia exponenciálnym rozdelením s parametrom $\lambda \in \mathbb{R}^+$, je náhodná premenná, ktorá má gama rozdelenie s parametrami $n \in \mathbb{N}$ a $\lambda \in \mathbb{R}^+$. Čiže pre $W_i \sim \text{Exp}(\lambda)$ platí:

$$\sum_{i=1}^{n} W_i = T_n \sim \text{Gama}(n, \lambda). \tag{5.9}$$

Popritom, ak parameter n je kladné celé číslo, tak hovoríme o špeciálnom prípade gama rozdelenia, ktoré sa nazýva Erlangovo rozdelenie pravdepodobnosti, dôsledkom čoho môžeme použiť aj zápis: $T_n \sim \text{Erlang}(n, \lambda)$. Medzi počtom nahlásených poistných nárokov $N(t)$ a dobami uplynutými medzi nahlásením dvoch po sebe idúcich poistných nárokov W_i platí nasledujúci vzťah:

$$N_t \geq n+1 \iff \sum_{i=1}^{n+1} W_i \leq t, \tag{5.10}$$

ktorý hovorí, že počet nahlásených poistných nárokov N_t za fixný čas $t > 0$ je väčší ako n práve vtedy, keď súčet $n+1$ prírastkov W_i je menší nanajvýš rovný času t [5], [13]. Spolu s využitím platnosti vzťahu (5.9) môžeme, tak ako aj v [5] a [13], písať:

$$\text{P}(N_t \geq n+1) = \text{P}\left(\sum_{i=1}^{n+1} W_i \leq t\right) = \text{P}(T_{n+1} \leq t) = F_{T_{n+1}}(t). \tag{5.11}$$

Vyššie uvedený vzťah (5.7) opisuje distribučnú funkciu Poissonovho rozdelenia pravdepodobnosti, ktorého použitím dostávame distribučnú funkciu náhodnej premennej T_{n+1}, ktorá sa riadi gama rozdelením pravdepodobnosti:

$$F_{T_{n+1}}(t) = \text{P}(N_t \geq n+1) = 1 - \text{P}(N_t \leq n) = 1 - \sum_{j=0}^{n} \frac{\text{e}^{-\lambda t}(\lambda t)^j}{j!}. \tag{5.12}$$

Prichádzame tak k záveru, že výsledky získané prostredníctvom gama rozdelenia v prípade časových bodov T_n a Poissonovho rozdelenia v prípade počtu poistných nárokov N_t by mali byť totožné, a zároveň aj k tomu, že na určenie rozdelenia pravdepodobnosti počtu nahlásených poistných nárokov N_t stačí určiť rozdelenie náhodnej premennej T_n [5]. Priblížme si ešte náhodný proces celkovej výšky nahlásených poistných nárokov, pre ktorý platí, že ak $\{N(t); t \geq 0\}$ je Poissonov proces, tak $\{S(t); t \geq 0\}$ je zložený Poissonov proces.

Definícia 5.2. [13] *(Zložený Poissonov proces)*
Náhodný proces $\{S(t); t \geq 0\}$ definovaný predpisom

$$S(t) = \sum_{i=1}^{N(t)} X_i,$$

sa nazýva zložený Poissonov proces s parametrom λ a distribučnou funkciou $F(\cdot)$, ak platí, že:

1. $\{N(t); t \geq 0\}$ je Poissonov proces s parametrom λ,
2. $\{X_i; i \geq 1\}$ je postupnosť nezávislých, rovnako rozdelených náhodných premenných s distribučnou funkciou $F(\cdot)$,
3. $\{X_i; i \geq 1\}$ a $\{N(t); t \geq 0\}$ sú navzájom nezávislé.

Zároveň platí, že ak $\{S(t); t \geq 0\}$ je zložený Poissonov proces, tak pre konkrétne hodnoty $t \geq 0$ má celková výška poistných nárokov zložené Poissonovo rozdelenie, ktoré označujeme zápisom:

$$S_t \sim \text{CoPoisson}(\lambda t, F_X), \tag{5.13}$$

kde F_X reprezentuje distribučnú funkciu rozdelenia pravdepodobnosti, podľa ktorého sa riadia výšky nahlásených poistných nárokov X_i, pre $i = 1, 2, \ldots$ Navyše vďaka vzťahom nazývaným ako Waldove identity pre zložené procesy, je možné vyjadriť aj strednú hodnotu a disperziu náhodného procesu $S(t)$ v ľubovoľnom fixnom čase t. Predpokladajme existenciu konečných momentov pre náhodné premenné $X_i \equiv X$ a náhodný proces $\{N(t); t \geq 0\}$ nezávislý od X_i pre $i = 1, 2, \ldots$ Potom platia nasledujúce vzťahy [13]:

$$\mathrm{E}(S(t)) = \mathrm{E}(X)\mathrm{E}(N(t)), \tag{5.14}$$

$$\mathrm{D}(S(t)) = \mathrm{D}(X)\mathrm{E}(N(t)) + \mathrm{E}^2(X)\mathrm{D}(N(t)). \tag{5.15}$$

V prípade, že náhodný proces $\{N(t); t \geq 0\}$ je Poissonov proces s parametrom $\lambda \in \mathbb{R}^+$, čiže rozdelenie pravdepodobnosti jednotlivých zložiek náhodného procesu je Poissonovo s parametrom λt, tak pre strednú hodnotu a disperziu platí: $\mathrm{E}(N(t)) = \mathrm{D}(N(t)) = \lambda t$. Waldove identity pre zložené procesy majú tak v tomto konkrétnom prípade tvar [13]:

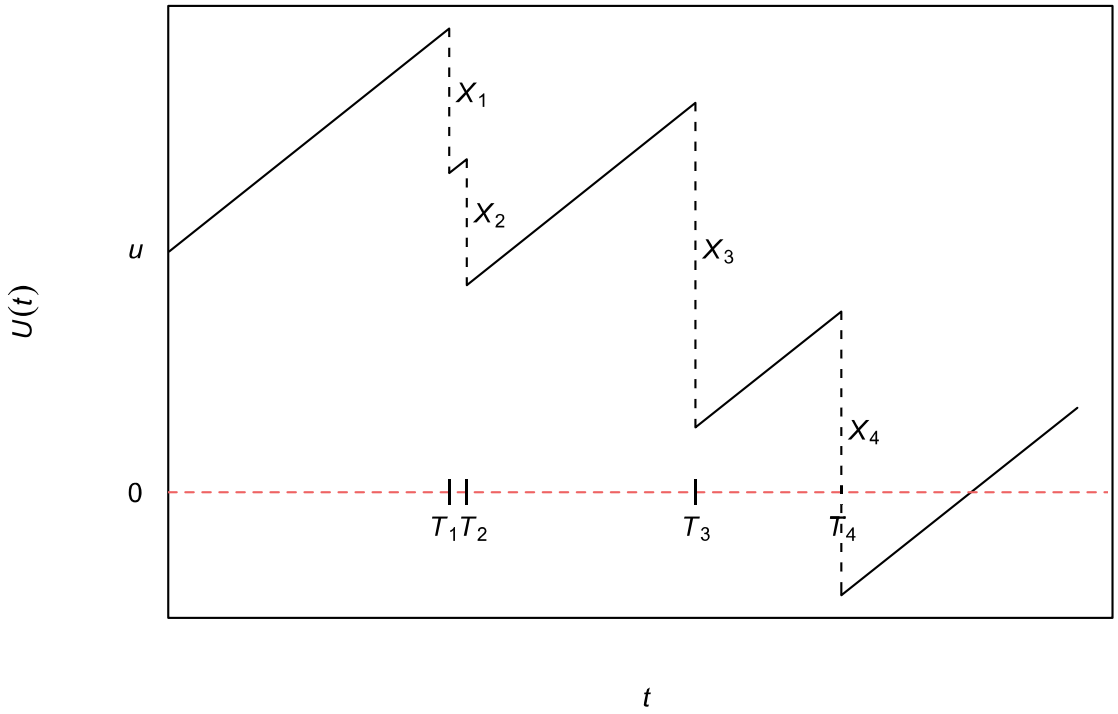
$$\mathrm{E}(S(t)) = \mathrm{E}(X)\lambda t, \tag{5.16}$$

$$\mathrm{D}(S(t)) = \mathrm{D}(X)\lambda t + \mathrm{E}^2(X)\lambda t = \left(\mathrm{D}(X) + \mathrm{E}^2(X)\right)\lambda t = \mathrm{E}\left(X^2\right)\lambda t. \tag{5.17}$$

Dodatočne sa ešte konkrétnejšie pozrime na výšku celkového poistného $c > 0$. Existuje niekoľko princípov výpočtu poistného, kde najjednoduchším z nich je princíp čistého poistného (*net premium principle*), ktorý je rovný strednej hodnote celkovej výšky poistných nárokov. V teórii ruinovania sa však zvykne používať poistný princíp očakávanej hodnoty (*expected value premium principle*), ktorý je pre akékoľvek $\theta > 0$, známe ako bezpečnostný koeficient či riziková prirážka, súčtom čistého poistného a rizikovej prirážky z čistého poistného, t. j.:

$$\begin{aligned} c &= \mathrm{E}(S_1) + \theta\mathrm{E}(S_1) = (1+\theta)\mathrm{E}(S_1) = (1+\theta)\mathrm{E}(N_1)\mathrm{E}(X), \\ c &= (1+\theta)\lambda\mathrm{E}(X). \end{aligned} \tag{5.18}$$

Ako sme už vyššie viackrát naznačili, podstatou teórie ruinovania je štúdium náhodného procesu prebytku poisťovne $\{U(t); t \geq 0\}$. Tento proces je matematicky popísaný modelom so spojitým časom, ktorého predpis je definovaný rovnicou (5.4) a tiež je graficky znázornený na Obr. 2.



Obr. 2: Grafické znázornenie realizácie náhodného procesu prebytku poisťovne (*zdroj: vlastné spracovanie v softvéri R* [9])

Primárnym cieľom teórie ruinovania je však aproximácia pravdepodobnosti zruinovania poisťovne, kde k zruinovaniu poisťovne dochádza v okamihu, keď je hodnota prebytku poisťovne prvýkrát záporná, teda vtedy, ak platí:

$$U(u,t) < 0, \quad \text{pre nejaké } t \geq 0. \tag{5.19}$$

Zruinovanie poisťovne môžeme sledovať v konečnom alebo nekonečnom časovom horizonte. V našom prípade budeme uvažovať nekonečný časový horizont, teda $t \in \langle 0, \infty)$. Pravdepodobnosť zruinovania poisťovne v závislosti od počiatočnej hodnoty prebytku potom definujeme zápisom:

$$\Psi(u) = \mathrm{P}\left(U(u,t) < 0, \text{ pre všetky } t \geq 0\right). \tag{5.20}$$

Na Obr. 2 pozorujeme graf znázorňujúci vývoj prebytku poisťovne, kde začínajúc s počiatočnou hodnotou prebytku poisťovne u sa táto hodnota lineárne zvyšuje v dôsledku prijatého poistného, a to až do časového bodu T_1, kedy dochádza k vyplateniu prvého poistného nároku vo výške X_1. Následne sa hodnota prebytku poisťovne opäť zvyšuje vďaka prijatému poistnému, no len do časového bodu T_2, kedy dôjde k ďalšiemu vyplateniu poistného nároku vo výške X_2. Zlomovým je časový bod T_4, v ktorom sa hodnota prebytku dostane prvýkrát pod nulovú hodnotu, čiže dôjde k zruinovaniu poisťovne a čas, v ktorom sa to stane, nazývame ako čas zruinovania poisťovne, ktorý je definovaný ako:

$$T(u) = \min_{t \geq 0}\{t : U(u,t) < 0\}. \tag{5.21}$$

Pravdepodobnosť zruinovania je možné vyjadriť aj prostredníctvom času zruinovania poisťovne $T(u)$, pretože k zruinovaniu poisťovne v nekonečnom časovom horizonte dôjde jedine vtedy, keď čas zruinovania poisťovne bude mať konečnú hodnotu [13]:

$$\Psi(u) = \mathrm{P}\left(T(u) < \infty\right). \tag{5.22}$$

Významným koeficientom používaným pri aproximácii pravdepodobnosti ruinovania je vyrovnávací koeficient R, nazývaný niekedy aj adjustačný koeficient, ktorý reprezentuje mieru rizika v procese prebytku poisťovne. V klasickom modeli teórie ruinovania, za predpokladu, že celková výška poistného je určená rovnicou (5.18), je vyrovnávací

koeficient jediné kladné riešenie rovnice:

$$M_X(R) = 1 + R(1+\theta)\mathrm{E}(X), \tag{5.23}$$

kde $M_X(\cdot)$ je momentová vytvárajúca funkcia náhodnej premennej X. Vzhľadom na to, že rovnica (5.23) je zadaná v implicitnom tvare, tak výpočet koreňov rovnice nemusí byť priamočiary, a preto sa riešenie často zvykne len ohraničiť. Jednou z najznámejších horných ohraničení pravdepodobnosti zruinovania poisťovne je Lundbergova nerovnosť.

Veta 5.1. [8] *(Lundbergova nerovnosť)*
Uvažujme klasický model teórie ruinovania, v ktorom nech platia všetky zavedené predpoklady. Nech výška poistného prijatého za strednú dobu čakania je vyššia ako stredná hodnota výšky poistných nárokov, t. j. nech

$$c\mathrm{E}(W_1) > \mathrm{E}(X_1).$$

Zároveň nech existuje vyrovnávací koeficient R. Potom pre pravdepodobnosť zruinovania poisťovne v nekonečnom časovom horizonte v prípade, že $u > 0$ platí:

$$\Psi(u) \leq \mathrm{e}^{-Ru}. \tag{5.24}$$

Dôkaz. Dôkaz je možné nájsť v publikácii [8]. □

V prípade, že sa v zloženom Poissonovom procese $\{S(t); t \geq 0\}$ riadia výšky nahlásených poistných nárokov exponenciálnym rozdelením pravdepodobnosti, tak vieme vyrovnávací koeficient z rovnice (5.23) vypočítať aj explicitne. Nech $X \sim \mathrm{Exp}\left(\frac{1}{\mu}\right)$ a stredná hodnota výšky poistných nárokov je μ. Momentová vytvárajúca funkcia exponenciálneho rozdelenia je potom podľa (2.6):

$$M_X(R) = \mathrm{E}\left(\mathrm{e}^{RX}\right) = \frac{1}{\mu}\int_0^\infty \mathrm{e}^{Rx}\mathrm{e}^{-\frac{1}{\mu}x}\mathrm{d}x = \cdots = \frac{1}{1-\mu R},$$

pričom platí, že vyrovnávací koeficient $R < 1/\mu$ (konečnosť integrálu). Po dosadení momentovej vytvárajúcej funkcie do rovnice (5.23) dostávame novú rovnicu:

$$\frac{1}{1-\mu R} = 1 + R(1+\theta)\mu, \tag{5.25}$$

ktorej riešením získame dva korene: prvým triviálnym riešením je $R = 0$ a druhým riešením je jediný kladný koreň rovný:

$$R = \frac{\theta}{(1+\theta)\mu}. \tag{5.26}$$

Veta 5.2. [6] *(Pravdepodobnosť ruinovania pre exponenciálny model)*
Uvažujme klasický model teórie ruinovania, v ktorom nech platia všetky zavedené predpoklady. Nech sa výšky nahlásených poistných nárokov riadia exponenciálnym rozdelením pravdepodobnosti s parametrom $1/\mu$ a nech výške celkového poistného zodpovedá $c = (1+\theta)\lambda\mu$, pričom $\theta > 0$. Potom pravdepodobnosť zruinovania poisťovne vzhľadom na počiatočný kapitál u je:

$$\Psi(u) = \frac{1}{1+\theta}\mathrm{e}^{-Ru},$$

kde R je vyrovnávací koeficient vyjadrený vzťahom (5.26).

Dôkaz. Dôkaz je možné nájsť v publikácii [6]. □

5.3 Tvorba aktuárskeho modelu v teórii ruinovania

V predchádzajúcich podkapitolách sme si predstavili model kolektívneho rizika a základy teórie ruinovania. Tie teraz využijeme pri tvorbe aktuárskeho modelu, kde si opäť predvedieme jednotlivé kroky postupu na ilustratívnom príklade. Cieľom poisťovní je nastaviť výšku počiatočného prebytku či výšku poistného tak, aby zabránila jej zruinovaniu. Poisťovňa v závislosti od nastavených parametrov a dostupných dát o výškach poistných nárokov dokáže odhadnúť pravdepodobnosť jej zruinovania. Priblížme si teda celý postup na nasledujúcom príklade.

Príklad 5.1. (*aktuárske modelovanie procesu rizika v teórii ruinovania*)

Uvažujme poisťovňu, ktorá poskytuje svojim klientom poistenie majetku, a kryje tak riziká súvisiace napríklad s požiarom, vytopením, krádežou, vandalizmom a podobne. Portfólio poistných zmlúv je homogénne (zmluvy sú podobné) a jeho veľkosť sa výrazne nemení. Poisťovňa má k dispozícii zoznam výšok $n = 600$ poistných nárokov nahlásených za posledné 3 mesiace spolu s časom, kedy boli poistné nároky nahlásené. Predpokladáme pritom, že tok poistných nárokov je možné modelovať pomocou Poissonovho procesu. Časť zoznamu o časoch a výškach nahlásených poistných nárokov je zobrazený v Tabuľke 8. Výšky nahlásených poistných nárokov sú evidované v eurách a celý dátový súbor je opäť k dispozícii na adrese: `http://www.iam.fmph.uniba.sk/ospm/data/05_data.txt`. Tiež uvažujeme, že jedna časová jednotka zodpovedá jednej

hodine. Na výpočet celkového poistného poisťovňa používa poistný princíp očakávanej hodnoty s bezpečnostným koeficientom $\theta = 0{,}20$. Hodnotu počiatočného prebytku poisťovňa nastavila za pomoci Lundbergovej nerovnosti na úroveň 10000 eur. Taktiež vieme, že poisťovňa používa klasický model teórie ruinovania. Cieľom poisťovne je pomocou získaných dát modelovať hodnotu prebytku poisťovne a aproximovať pravdepodobnosť jej zruinovania.

Tabuľka 8: Ukážka dátového súboru o výškach poistných nárokov nahlásených za posledné 3 mesiace spolu s časom ich nahlásenia.

poradové číslo i	čas nahlásenia T_i	výška poistného nároku X_i
1	1,45	1467,96
2	3,49	2312,27
3	4,18	1027,50
4	6,58	818,64
5	7,92	1344,80
⋮	⋮	⋮
598	2286,52	1385,77
599	2295,71	1820,87
600	2297,31	1463,16

(*zdroj*: *vlastné spracovanie*)

Riešenie.

Na začiatok je opäť dobré položiť si otázku, čo chceme modelovať a prečo to chceme modelovať. Odpoveď je jasná. Na základe získaných dát chceme modelovať hodnotu prebytku poisťovne a pravdepodobnosť jej zruinovania, a to napríklad kvôli tomu, aby poisťovňa vedela v budúcnosti lepšie nastaviť počiatočnú hodnotu prebytku u, prípadne výšku poistného c, aby tak minimalizovala pravdepodobnosť zruinovania poisťovne. Pozrime sa najskôr na dáta, ktoré máme k dispozícii a nasledujme postupnosť krokov tvorby aktuárskeho modelu, podobne ako v Príklade 4.1.

Krok 1: získanie, načítanie a analýza dát + využitie aktuárskych znalostí

Po načítaní dát v softvéri R sme zistili, že dostupné dáta sú zoradené vzostupne podľa času T_i, v ktorom došlo k nahláseniu poistného nároku. Výšky nahlásených poistných nárokov X_i sú všetky nenulové a v dátach sa nenachádza žiadna chýbajúca hodnota. Ako sme už uviedli vyššie, chceme modelovať hodnotu prebytku poisťovne, ktorá je definovaná vzťahom (5.4). Aby sme tak vedeli urobiť, musíme byť schopní modelovať aj celkovú výšku poistných nárokov $S(t)$ definovanú vzťahom (5.5), z čoho vyplýva, že musíme vedieť modelovať aj výšky nahlásených poistných nárokov X_i a tok poistných nárokov $N(t)$. Zo zadania príkladu však vieme, že tok poistných nárokov je možné modelovať Poissonovým procesom, čiže na základe definície Poissonovho procesu vieme povedať, že doby uplynuté medzi nahlásením dvoch po sebe nasledujúcich poistných nárokov W_i by sa mali riadiť exponenciálnym rozdelením pravdepodobnosti s parametrom λ. Parameter λ nazývame aj intenzita príchodov poistných nárokov, a ak by sme poznali jeho hodnotu, vedeli by sme povedať, ako často dochádza k nahláseniu poistného nároku. Bolo by teda vhodné odhadnúť hodnotu parametra λ, a zároveň tak overíme, či sa náhodné premenné W_i skutočne riadia exponenciálnym rozdelením pravdepodobnosti. Dostali sme sa teda k tomu, že hodnotu prebytku poisťovne $U(t)$ budeme v tomto prípade schopní modelovať vtedy, ak budeme vedieť modelovať náhodné premenné W_i a X_i.

Modelovanie doby uplynutej medzi nahlásením dvoch po sebe nasledujúcich poistných nárokov W_i

Dáta z Tabuľky 8 nám poskytujú informácie o výškach nahlásených poistných nárokov spolu s časom kedy boli nároky nahlásené. Doby uplynuté medzi nahlásením dvoch po sebe idúcich poistných nárokov priamo nepoznáme, no vieme ich vďaka platnosti vzťahu (5.8) a predpokladu, že $W_1 = T_1$ jednoducho dopočítať. Keď už máme postupnosť náhodných premenných W_i k dispozícii, tak môžeme pokračovať ďalším krokom tvorby aktuárskeho modelu.

Krok 2: voľba konkrétneho matematického modelu

Definícia Poissonovho procesu hovorí, že ak sa doby uplynuté medzi nahlásením po sebe nasledujúcich poistných nárokov W_i riadia exponenciálnym rozdelením pravdepodobnosti s *rate* parametrom $\lambda \in \mathbb{R}^+$, tak náhodný proces počtu poistných nárokov

je Poissonov proces s parametrom λ. V zadaní nášho príkladu sa predpokladá, že tok poistných nárokov je možné modelovať Poissonovým procesom, z čoho vyplýva, že náhodné premenné W_i by sa mali riadiť exponenciálnym rozdelením pravdepodobnosti. K modelovaniu náhodnej premennej W_i tak budeme pristupovať parametricky a overíme, či sa doby uplynuté medzi nahlásením dvoch po sebe idúcich poistných nárokov W_i naozaj riadia exponenciálnym rozdelením. Voľba matematického modelu je teda zrejmá, zvolíme exponenciálne rozdelenie pravdepodobnosti.

Krok 3: odhadovanie parametrov zvoleného modelu

Predpokladáme teda, že $W_i \sim \text{Exp}(\lambda)$, kde $\lambda \in \mathbb{R}^+$ je *rate* parameter. Parameter λ odhadujeme rovnako ako v Príklade 4.1 v prostredí štatistického softvéru R pomocou funkcie `fitdist` z balíka `fitdistrplus` [2], pričom opäť volíme metódu maximálnej dobrej zhody založenú na Cramérovej-von Misesovej vzdialenosti. Hlavička funkcie `fitdist` má v tomto konkrétnom prípade tvar:

```
fitdist(W, distr="exp", method="mge", discrete=FALSE, gof="CvM").
```

Odhadom *rate* parametra λ je hodnota 0,2598. Pre náhodné premenné W_i tak môžeme písať: $W_i \sim \text{Exp}(0{,}2598)$.

Krok 4: validácia a overovanie vhodnosti modelu

V predchádzajúcom kroku sme odhadli hodnotu parametra exponenciálneho rozdelenia pravdepodobnosti. Je však potrebné overiť, či je exponenciálne rozdelenie dostatočne vhodné a presné. Pozrieme sa preto na p-hodnoty Kolmogorovovho-Smirnovovho testu, Cramérovho-von Misesovho testu a Andersonovho-Darlingovho testu. Taktiež sa pozrieme na hodnotu Akaikeho informačného kritéria, prípadne by sme sa mohli pozrieť na P-P graf či Q-Q graf. Na základe p-hodnôt, ktoré sú uvedené v Tabuľke 9, je možné konštatovať, že náhodné premenné W_i by sa mohli riadiť exponenciálnym rozdelením pravdepodobnosti.

Krok 5: vyhodnotenie výsledkov

Rozumné by bolo zvoliť aj iné rozdelenia pravdepodobnosti, aby sme mali výsledky s čím porovnať. Rozhodli sme sa pre porovnanie zvoliť niektoré z rozdelení, ktoré sú rovnako ako exponenciálne rozdelenie spojité a definované na množine nezáporných reálnych čísel. Takými sú napríklad log-normálne rozdelenie, gama rozdelenie

či Weibullovo rozdelenie. Vrátime sa tak ku kroku 2 a postup zopakujeme s každým rozdelením pravdepodobnosti zvlášť. Hodnoty odhadnutých parametrov všetkých zvolených rozdelení, hodnoty AIC a p-hodnoty testov dobrej zhody môžeme pozorovať v Tabuľke 9, pričom sa nám na základe p-hodnôt testov dobrej zhody a hodnoty AIC potvrdilo, že najvhodnejším rozdelením pre modelovanie dôb uplynutých medzi nahlásením dvoch po sebe idúcich poistných nárokov je exponenciálne rozdelenie pravdepodobnosti s parametrom $\lambda = 0{,}2598$.

Tabuľka 9: Zhrnutie odhadnutých parametrov pre doby uplynuté medzi nahlásením dvoch po sebe nasledujúcich poistných nárokov, hodnôt AIC a p-hodnôt testov dobrej zhody vybraných rozdelení pravdepodobnosti.

rozdelenie W_i	odhad parametrov		AIC	K-S test	C-vM test	A-D test
Exp(λ)	**0,2598**	-	**2813,0944**	**0,9643**	**0,8726**	**0,8855**
LN(μ, σ^2)	0,9048	$1{,}1865^2$	2920,5055	0,0211	0,0324	0,0004
Gama(α, β)	0,9631	0,2474	2815,6388	0,9787	0,9326	0,8954
Wei(κ, θ)	0,9790	3,8615	2815,5019	0,9758	0,9198	0,8867

(*zdroj*: *vlastné spracovanie*)

Odhadnutý parameter exponenciálneho rozdelenia $\lambda = 0{,}2598 \simeq \frac{1}{4}$, nazývaný aj intenzita príchodov poistných nárokov, môžeme interpretovať tak, že v priemere sa jeden poistný nárok nahlási približne raz za 4 hodiny.

Modelovanie výšky nahlásených poistných nárokov X_i

Za sledované obdobie bolo nahlásených 600 poistných nárokov, ktorých výšky sú známe. V prvom kroku tvorby aktuárskeho modelu sme sa pozreli na dáta konkrétnejšie a zistili sme, že všetky výšky sú nenulové. S modelovaním výšky nenulových poistných nárokov už máme skúsenosť z Príkladu 4.1, a preto budeme v tomto príklade stručnejší. Postupujme opäť podľa zavedených krokov tvorby aktuárskeho modelu.

Krok 2: voľba konkrétneho matematického modelu

K modelovaniu výšky nahlásených poistných nárokov budeme opäť pristupovať parametricky. Zvolíme parametrické rozdelenie pravdepodobnosti, ktoré by mohlo

čo najpresnejšie opisovať správanie výšok poistných nárokov. Z minulých skúseností už vieme, že na modelovanie nenulových výšok poistných nárokov sú vhodné spojité rozdelenia pravdepodobnosti definované na množine nezáporných reálnych čísel. Pozrime sa ešte na základné opisné charakteristiky pre výšky nahlásených poistných nárokov, ktoré sú uvedené v Tabuľke 10.

Tabuľka 10: Základné opisné charakteristiky výšok nahlásených poistných nárokov.

minimálna hodnota	maximálna hodnota	stredná hodnota	smerodajná odchýlka	koeficient šikmosti	koeficient excesu
451,06	3005,59	1493,52	430,02	0,26	−0,16

(*zdroj*: *vlastné spracovanie*)

Na základe koeficientu šikmosti, ktorý je v tomto prípade rovný 0,26, vieme povedať, že vhodnými rozdeleniami sú pozitívne zošikmené rozdelenia, čiže napríklad: exponenciálne, gama, log-normálne, Weibullovo či Burrovo rozdelenie pravdepodobnosti. Skúsme na modelovanie výšky poistných nárokov ako prvé zvoliť log-normálne rozdelenie pravdepodobnosti.

Krok 3: odhadovanie parametrov zvoleného modelu

Uvažujme, že $X_i \sim \text{LN}(\mu, \sigma^2)$, potom *log-scale* parameter $\mu \in \mathbb{R}$ a *log-shape* parameter $\sigma \in \mathbb{R}^+$ odhadneme v prostredí štatistického softvéru R pomocou funkcie `fitdist`, pričom aj v tomto prípade zvolíme metódu maximálnej dobrej zhody s účelovou funkciou Cramérovho-von Misesovho testu. Výsledkom sú hodnoty parametrov: $\mu = 7{,}2863$ a $\sigma = 0{,}2932$.

Krok 4: validácia a overovanie vhodnosti modelu

Parametre log-normálneho rozdelenia sa nám podarilo odhadnúť. Teraz musíme však znova overiť, či log-normálne rozdelenie pravdepodobnosti dostatočne presne opisuje výšky poistných nárokov. Rovnako ako v predchádzajúcich prípadoch validácie, tak aj teraz využijeme funkciu `gofstat` z balíku `fitdistrplus` [2], ktorej výstupom sú hodnoty štatistík Kolmogorovovho-Smirnovovho testu, Cramérovho-von Misesovho

testu a Andersonovho-Darlingovho testu, a tiež hodnota Akaikeho informačného kritéria. P-hodnoty týchto testových štatistík získame z výstupu funkcii `ks.test`, `cvm.test` a `ad.test`, ktoré nájdeme v balíku `goftest` [4]. P-hodnota Kolmogorovovho-Smirnovovho testu a Cramérovho-von Misesovho testu nadobúdajú hodnoty nad hladinou významnosti 5 %. Konkrétne výšky p-hodnôt sú uvedené v Tabuľke 11 a pre ich replikáciu je možné použiť programový kód, ktorý je uvedený v pasívnej prílohe tejto diplomovej práce. V Tabuľke 11 vidíme aj p-hodnotu Andersonovho-Darlingovho testu, ktorej hodnota je menej ako 5 %. Na základe p-hodnôt testov dobrej zhody tak môžeme tvrdiť, že log-normálne rozdelenie nie je vhodné na modelovanie výšok nahlásených poistných nárokov. Jedným z dôvodov prečo sa log-normálne rozdelenie pravdepodobnosti nejaví byť vhodným rozdelením môže byť aj to, že toto rozdelenie patrí k rozdeleniam s ťažkým chvostom. Avšak z Tabuľky 10 môžeme pozorovať, že sa v dátach nenachádza žiadna extrémne vysoká hodnota, a tak by bolo pri modelovaní výšky nahlásených poistných nárokov rozumnejšie použiť rozdelenia s ľahkým pravým chvostom. Je tak potrebné vrátiť sa ku kroku 2 a vybrať iné rozdelenie pravdepodobnosti.

Krok 5: vyhodnotenie výsledkov

Ako ďalšie modely sme zvolili gama rozdelenie pravdepodobnosti a Weibullovo rozdelenie pravdepodobnosti. Neuvádzame však znova celý postup kalibrácie a validácie modelov. Hodnoty odhadnutých parametrov všetkých troch zvolených rozdelení, hodnoty Akaikeho informačného kritéria a p-hodnoty testov dobrej zhody sú zaznamenané v Tabuľke 11.

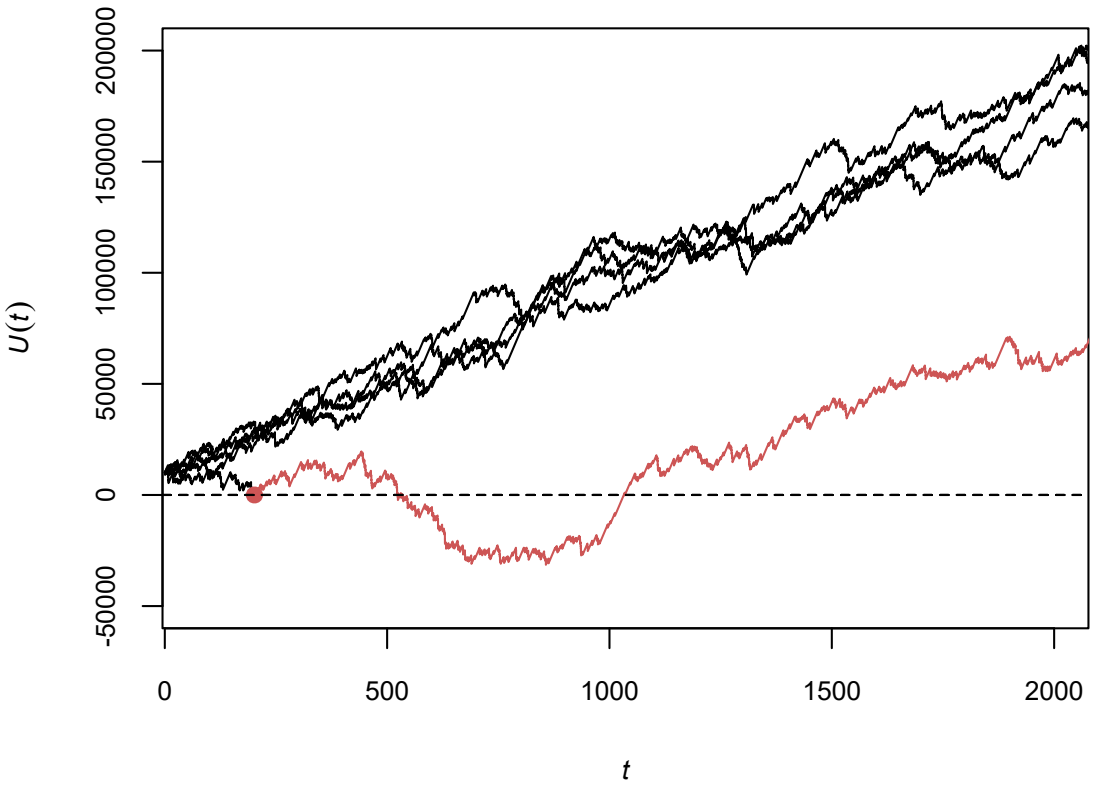
Tabuľka 11: Zhrnutie odhadnutých parametrov pre výšky poistných nárokov, hodnôt AIC a p-hodnôt testov dobrej zhody vybraných rozdelení pravdepodobnosti.

rozdelenie X_i	odhad parametrov		AIC	K-S test	C-vM test	A-D test
LN(μ, σ^2)	7,2863	$0{,}2932^2$	9013,1251	0,1910	0,3458	0,0378
Gama(α, β)	**11,8160**	**0,0078**	**8982,5070**	**0,6050**	**0,7486**	**0,4012**
Wei(κ, θ)	3,8920	1630,0109	8989,9342	0,6763	0,5707	0,2439

(*zdroj*: *vlastné spracovanie*)

Najvhodnejšie rozdelenie pre modelovanie výšky poistných nárokov sa na základe hodnoty AIC a p-hodnôt testov dobrej zhody javí byť gama rozdelenie pravdepodobnosti s parametrami: $\alpha = 11{,}8160$ a $\beta = 0{,}0078$.

Keďže predpokladáme, že tok poistných nárokov sa dá modelovať Poissonovým procesom s odhadnutým parametrom $\lambda = 0{,}2598 \simeq \frac{1}{4}$ a výšky nahlásených poistných nárokov sa riadia gama rozdelením pravdepodobnosti, pričom tok poistných nárokov a výšky poistných nárokov sú nezávislé, tak môžeme na základe Definície 5.2 tvrdiť, že náhodný proces $\{S(t); t \geq 0\}$ je zložený Poissonov proces s parametrom λ a distribučnou funkciou gama rozdelenia pravdepodobnosti. Po získaní všetkých týchto znalostí sme schopní modelovať hodnotu prebytku poisťovne, a to v našom prípade pomocou Monte Carlo simulácie. Ukážka simulácie hodnoty prebytku poisťovne je zobrazená na Obr. 3. Pre prehľadnosť a zrozumiteľnosť grafu sme zvolili len päť simulačných behov, pričom počiatočná hodnota prebytku je daná v zadaní príkladu a jej hodnota je 10000 eur.



Obr. 3: Vývoj hodnoty prebytku poisťovne v čase (*zdroj: vlastné spracovanie v softvéri R* [9])

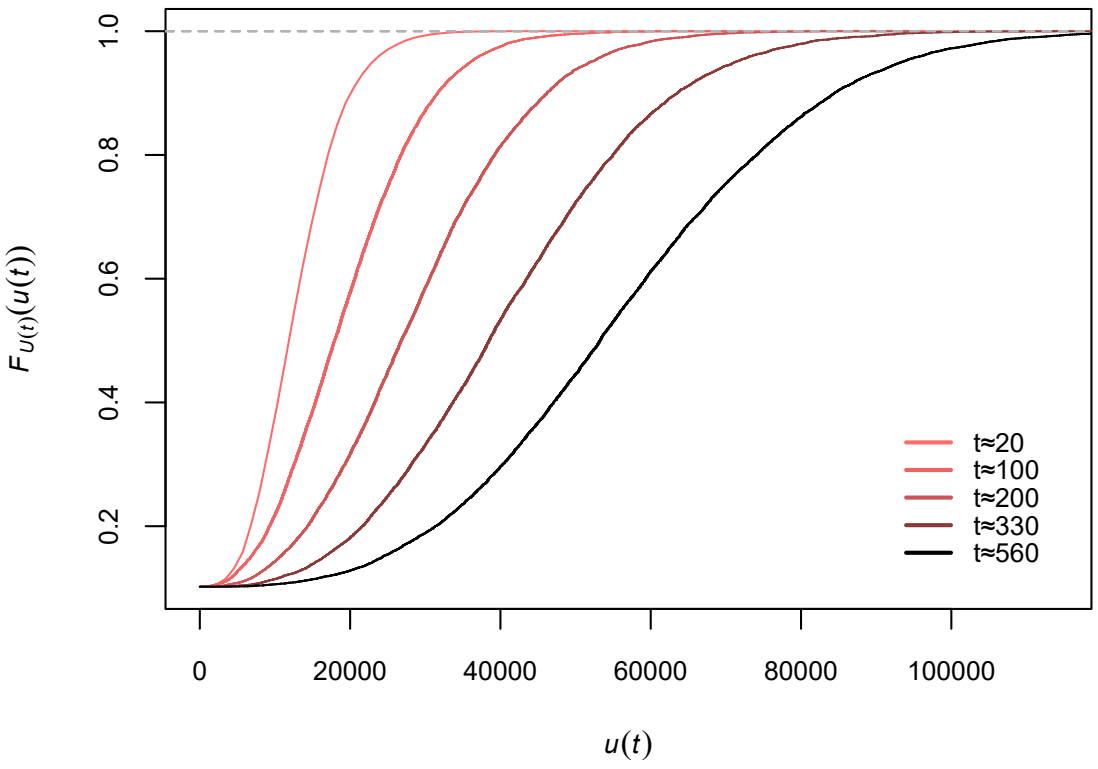
Na Obr. 3 tak môžeme pozorovať vývoj piatich rôznych procesov hodnoty prebytku poisťovne, kde štyri z nich sú znázornené čiernou farbou a jeden proces je od určitého bodu znázornený červenou farbou. Tento proces znázorňuje prípad, kedy došlo k zruinovaniu poisťovne a poisťovňa už nie je schopná plniť svoje záväzky voči poisteným klientom. Vyznačený bod zodpovedá času, v ktorom hodnota prebytku poisťovne prvýkrát klesla pod nulovú hodnotu. Po prvom takomto poklese už vývoj hodnoty prebytku poisťovne nesledujeme, a to aj v prípade, že by hodnota prebytku poisťovne bola vždy už len kladná.

Z Obr. 3 môžeme tiež pozorovať, že ak by sme zvýšili počiatočnú hodnotu prebytku, mohli by sme zabrániť zruinovaniu poisťovne. Prípadne by sme zruinovaniu mohli predísť aj tým, že zvýšime poisteným klientom poistné. Avšak tým riskujeme, že o klientov úplne prídeme. Zamerajme sa teda na to, ako bez akejkoľvek numerickej optimalizácie určiť primeranú hodnotu počiatočného prebytku poisťovne u. Je zrejmé, že čím je vyššia hodnota počiatočného prebytku, tým je nižšia pravdepodobnosť zruinovania poisťovne. Lenže poisťovňa si nemôže dovoliť vyhradiť príliš vysokú hodnotu počiatočného prebytku. Rovnako ani pravdepodobnosť zruinovania nebude nikdy nulová. Prichádzame preto k úsudku, že ak si určíme zmysluplnú hornú hranicu pravdepodobnosti zruinovania, tak na základe Lundbergovej nerovnosti definovanej vzťahom (5.24) budeme schopní dopočítať, ako určiť hodnotu počiatočného prebytku. Najprv však potrebujeme poznať hodnotu vyrovnávacieho koeficientu, ktorou je jediné kladné riešenie rovnice (5.23), pričom $M_X(\cdot)$ je v našom prípade momentová vytvárajúca funkcia gama rozdelenia pravdepodobnosti. Riešenie rovnice (5.23) zadanej v implicitnom tvare získame výpočtom v prostredí štatistického softvéru R. Výsledkom je hodnota vyrovnávacieho koeficientu $R = 0{,}00021439$. Ak chceme teda zistiť, aká má byť hodnota počiatočného prebytku pri hornej hranici zruinovania napríklad 5 %, tak použijeme Lundbergovu nerovnosť a získame výsledok:

$$\begin{aligned} \mathrm{e}^{-Ru} &\leq 0{,}05, \\ u &\geq -\frac{\ln 0{,}05}{R} = 13973{,}28 \text{ eura}. \end{aligned}$$

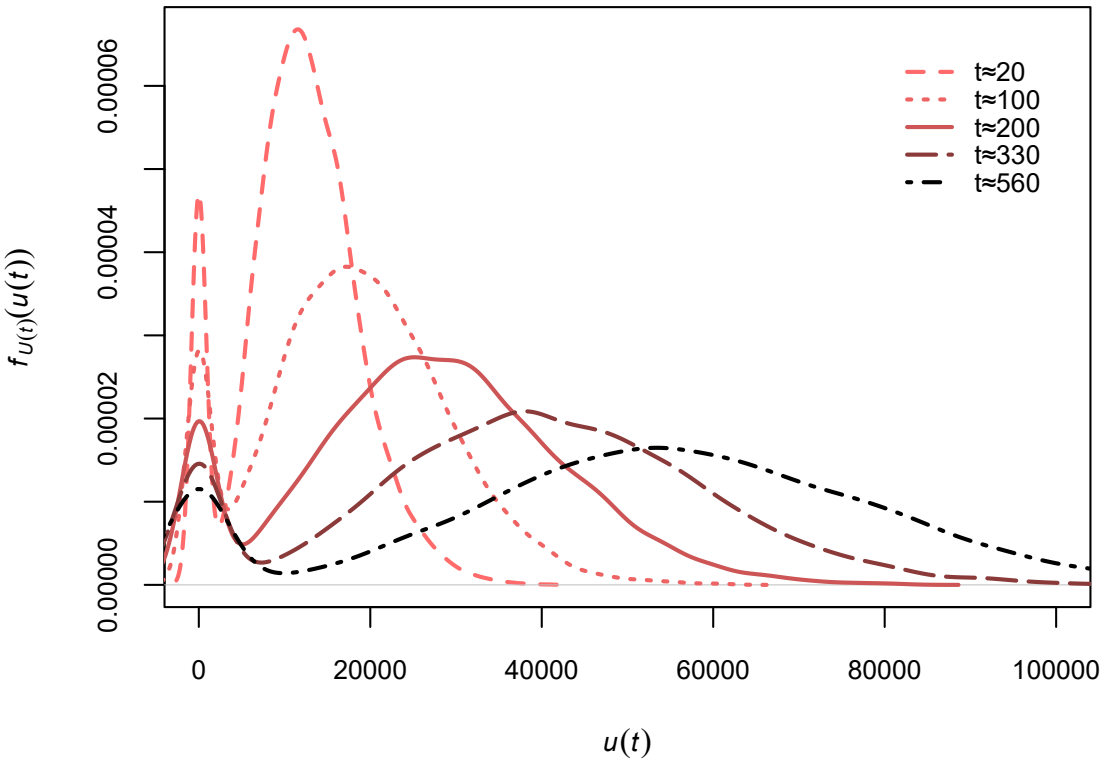
V zadaní príkladu je hodnota počiatočného prebytku poisťovne rovná 10000 eur, a tak sa zdá, že pravdepodobnosť zruinovania v nekonečnom časovom horizonte bude vyššia ako 5 %.

Pozrime sa na distribučnú funkciu hodnoty prebytku poisťovne, ktorá je zobrazená na Obr. 4, a to v rôznych časoch t. Ide o výsledok jednej konkrétnej Monte Carlo simulácie, pričom počet simulačných behov je 10000. Doplníme, že ak dôjde k zruinovaniu poisťovne, tak hodnotu prebytku poisťovne považujeme za nulovú a jej ďalší vývoj už neuvažujeme. Rovnako neberieme do úvahy ani žiadnu formu penalizácie v prípade zruinovania poisťovne. Z Obr. 4 môžeme tiež pozorovať, že s rastúcim časom rastie aj hodnota prebytku poisťovne, čomu nasvedčujú aj čierne postupne rastúce krivky z Obr. 3. Nakoniec skúsme objasniť význam funkčnej hodnoty pre rôzne distribučné funkcie. Funkčná hodnota distribučnej funkcie hodnoty prebytku poisťovne vyjadruje pravdepodobnosť, s akou hodnota prebytku poisťovne bude nanajvýš napríklad 40000 eur. A tak je zrejmé, že v dlhšom časovom intervale je pravdepodobnosť dosiahnutia maximálne tejto hranice nižšia, keďže poisťovňa môže zarábať dlhšie.



Obr. 4: Distribučná funkcia hodnoty prebytku poisťovne vo vybraných časoch, ktorú je možné získať pomocou simulácie (*zdroj: vlastné spracovanie v softvéri R* [9])

K vyššie uvedenej simulovanej distribučnej funkcii je na Obr. 5 znázornená aj prislúchajúca simulovaná funkcia hustoty hodnoty prebytku poisťovne.



Obr. 5: Simulovaná funkcia hustoty hodnoty prebytku poisťovne vo vybraných časoch (*zdroj: vlastné spracovanie v softvéri R* [9])

Na záver sa vráťme k už mnohokrát spomínanej pravdepodobnosti zruinovania poisťovne. Načrtli sme, že pravdepodobnosť zruinovania poisťovne, ktorej počiatočný kapitál je 10000 eur, výšky poistných nárokov sa riadia gama rozdelením pravdepodobnosti s parametrami $\alpha = 11{,}8160$ a $\beta = 0{,}0078$, a jeden poistný nárok sa nahlási v priemere približne raz za štyri časové jednotky, by mala byť vyššia ako 5 %. Priemernú pravdepodobnosť zruinovania poisťovne sme odhadli v softvéri R pomocou 1000 Monte Carlo simulácií, kde každá z nich bola nastavená na 1000 simulačných behov. Zistili sme, že očakávaná pravdepodobnosť zruinovania poisťovne je na úrovni 10,32 %. Zároveň sme sa pozreli aj na priemerný čas zruinovania poisťovne, ktorého hodnota je 126,83 hodín. Rovnako by nás mohlo zaujímať napríklad to, po ktorom nároku priemerne dôjde k zruinovaniu poisťovne. Odpoveďou je, že sa tak stane po približne 45. poistnom nároku, pričom všetkých nahlásených poistných nárokov je 600. Upresníme ešte, že v každom

simulačnom behu bolo vygenerovaných vždy 600 nenulových výšok poistných nárokov tak, aby simulácia zodpovedala počtu nahlásených poistných nárokov v Príklade 5.1. Konečný čas tak nie je pevne stanovený, a preto sa môže v každom simulačnom behu trochu líšiť od konečného času uvedeného v Príklade 5.1. Aj napriek tomu však môžeme konštatovať, že ak dôjde k zruinovaniu poisťovne, tak to bude v pomerne krátkom časovom intervale. Doplníme ešte, že v pasívnej prílohe tejto diplomovej práce sa nachádza programový kód k prezentovanému príkladu, a teda je možné pozrieť sa na modelovanie hodnoty prebytku poisťovne a pravdepodobnosti jej zruinovania aj pre iné hodnoty parametrov.

Záver

Matematické modelovanie je jednou z hlavných oblastí záujmu aktuárskej vedy, kde na základe techník z teórie pravdepodobnosti, matematickej štatistiky, finančnej matematiky či demografickej štatistiky je možné skonštruovať efektívne a dostatočne presné aktuárske modely využívané v mnohých okruhoch matematického modelovania. Cieľom tejto diplomovej práce bolo preto prehľadne a zrozumiteľne predstaviť jednotlivé kroky postupu tvorby aktuárskych modelov a uviesť vybrané okruhy matematického modelovania do kontextu tvorby aktuárskych modelov.

Prvá kapitola diplomovej práce predstavuje všeobecný úvod, v ktorom sme objasnili, čo je aktuárska veda a aktuárske modelovanie, a následne sme uviedli príklady využitia matematického modelovania v poisťovníctve, a to v životnom, neživotnom i zdravotnom poistení. V závere úvodnej kapitoly sme sa zamerali na prehľad existujúcej odbornej literatúry, z ktorej sme pri písaní diplomovej práce vychádzali.

V druhej kapitole sme sa zamerali na niektoré poznatky z teórie pravdepodobnosti a na štatistické metódy využívané v oblasti aktuárstva. Začali sme podkapitolou 2.1, kde sme sa venovali základným pojmom a definíciám z teórie pravdepodobnosti a matematickej štatistiky. V podkapitolách 2.2 a 2.3 sme definovali vybrané spojité a diskrétne rozdelenia pravdepodobnosti, ku ktorým sme tiež zaviedli potrebné matematické označenia a skratky, ktoré sme v ďalších kapitolách diplomovej práce často využívali. Nakoniec sme sa v podkapitole 2.4 venovali vybraným štatistickým metódam, ktoré sa používajú pri konštruovaní a validácii aktuárskych modelov. Medzi tieto štatistické metódy radíme niektoré známe testy dobrej zhody, a to konkrétne Kolmogorovov-Smirnovov test, Cramérov-von Misesov test a Andersonov-Darlingov test.

Následne sme v tretej kapitole predstavili postupnosť krokov používaných pri tvorbe aktuárskeho modelu a oboznámili sme sa s niektorými tematickými okruhmi matematického modelovania v oblasti aktuárstva. V prvej podkapitole 3.1 sme sa venovali aktuárskemu modelovaniu v demografii, a to konkrétne modelovaniu úmrtnostných kriviek. V nasledujúcich podkapitolách sme sa zaoberali aj aktuárskym modelovaním výnosových kriviek, aktuárskym modelovaním celkovej výšky poistných nárokov v modeli kolektívneho rizika a nakoniec sme si priblížili aktuárske modelovanie vývojových trojuholníkových schém. Pri každej oblasti modelovania sme sa zamysleli a odpovedali

na otázku, čo chceme modelovať a prečo to chceme modelovať. Nezaoberali sme sa však tým, ako by to bolo vhodné modelovať. Pozornosť sme tejto otázke venovali pri modelovaní celkovej výšky poistných nárokov v modeli individuálneho rizika vo štvrtej kapitole a pri modelovaní procesu rizika v teórii ruinovania v piatej kapitole.

Najskôr sme v podkapitole 4.1 opísali model individuálneho rizika, ktorý je definovaný vzťahom (4.1), spolu s jeho predpokladmi a základnými charakteristikami. Následne sme pokračovali tvorbou aktuárskeho modelu. V tomto prípade sme sa zamerali najmä na otázku, ako to chceme modelovať. K tejto otázke sa viaže celá postupnosť krokov tvorby aktuárskeho modelu, ktorú sme prehľadne ilustrovali na riešení Príkladu 4.1. Cieľom bolo zostrojiť aktuársky model, na základe ktorého by bolo možné čo najpresnejšie odhadovať celkovú výšku poistných nárokov. Výsledkom modelovania je Obr. 1, ktorý znázorňuje porovnanie normálnej aproximácie hustoty a simulovanej funkcie hustoty celkovej výšky poistných nárokov a viaže sa k nemu aj slovné vyhodnotenie získaných výsledkov.

V poslednej kapitole tejto diplomovej práce sme sa zaoberali aktuárskym modelovaním procesu rizika v teórii ruinovania. V prvom rade sme sa v podkapitole 5.1 venovali predstaveniu modelu kolektívneho rizika definovaného vzťahom (5.1), keďže zovšeobecnenie tohto modelu tvorí základ teórie ruinovania. Klasický model teórie ruinovania je následne definovaný vzťahom (5.4) a celý je opísaný v podkapitole 5.2. Posledná podkapitola tejto diplomovej práce je venovaná druhému ilustratívnemu Príkladu 5.1, na ktorom sme demonštrovali postup tvorby aktuárskeho modelu v teórii ruinovania, kde sme opäť priniesli aj grafické výstupy. Cieľom bolo pomocou získaných dát modelovať hodnotu prebytku poisťovne a aproximovať pravdepodobnosť jej zruinovania.

Hlavným prínosom tejto diplomovej práce je prehľadné spracovanie postupu tvorby aktuárskeho modelu v modeli individuálneho rizika a v teórii ruinovania. Výsledky príkladov je možné replikovať, keďže v pasívnej prílohe diplomovej práce k ním uvádzame programové kódy vytvorené v prostredí štatistického softvéru R. Možné rozšírenie diplomovej práce by mohlo predstavovať pridanie nadstavbových krokov tvorby aktuárskeho modelu, ako napríklad úprava modelu na budúce obdobia, prípadne spracovanie procesu tvorby aktuárskeho modelu aj v ďalších okruhoch matematického modelovania.

Zoznam použitej literatúry

[1] Anděl, J.: *Základy matematické statistiky*. Prvé vydanie, Matfyzpress, Praha, 2005, ISBN 80-86732-40-1.

[2] Delignette-Muller, M. L., Dutang, C. [online]: *fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data*. R package version 1.1-8, 2022 [cit. 11.04.2023] Dostupné na adrese: `https://cran.r-project.org/web/packages/fitdistrplus/fitdistrplus.pdf`.

[3] Dickson, D. C. M.: *Insurance Risk and Ruin*. First Edition, Cambridge University Press, New York, 2005, ISBN 0-521-84640-4.

[4] Faraway, J., Marsaglia, G., Marsaglia, J., Baddeley, A. [online]: *goftest: Classical Goodness-of-Fit Tests for Univariate Distributions*. R package version 1.2-3, 2021 [cit. 11.04.2023] Dostupné na adrese: `https://cran.r-project.org/web/packages/goftest/goftest.pdf`.

[5] Horáková, G., Páleš, M., Slaninka, F.: *Teória rizika v poistení*. Prvé vydanie, Wolters Kluver, Bratislava, 2015, ISBN 978-8-081-68273-5.

[6] Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M.: *Modern Actuarial Risk Theory Using R*. Second Edition, Springer-Verlag Berlin Heidelberg, 2008, ISBN 978-3-642-03407-7.

[7] Klugman, S. A., Panjer, H. H., Willmot, G. E.: *Loss Models: From Data to Decisions*. Fifth Edition, John Wiley & Sons Ltd, New Jersey, 2019, ISBN 978-1-119-52373-4.

[8] Mikosch, T.: *Non-Life Insurance Mathematics*. Springer, University of Copenhagen, Copenhagen, 2006, ISBN 10 3-540-40650-6.

[9] R Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. [cit. 14.01.2023] Dostupné na adrese: `https://www.R-project.org/`.

[10] Rejková, A.: *Statistické metody v pojistné matematice*. Bakalárska práca, Prírodovedecká fakulta, Masarykova univerzita, Brno, 2015.

[11] Rényi, A.: *Teorie pravděpodobnosti*. Prvé vydanie, Academia, Praha, 1972.

[12] Szűcs, G.: *Matematické modelovanie v poisťovníctve*. Skriptá k predmetu Poisťovníctvo, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave, Bratislava, 2022.

[13] Szűcs, G.: *Pravdepodobnostné modely v poisťovníctve*. Skriptá k predmetu Poistná matematika, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave, Bratislava, 2022.

[14] Szűcs, G.: *Prednášky z demografickej štatistiky*. Skriptá k predmetu Demografická štatistika, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave, Bratislava, 2021.

[15] Ventceľová, J. S.: *Teória pravdepodobnosti*. Prvé vydanie, Alfa, Bratislava, 1973.

Prílohy

Príloha A.

Programový kód k Príkladu 4.1 vytvorený v prostredí štatistického softvéru R.

```
# K spusteniu tohto programového kódu je potrebné mať funkčné
# pripojenie k internetu.

# inštalácia potrebných balíkov
install.packages(c("fitdistrplus", "goftest", "e1071"))

# načítanie balíkov
require(fitdistrplus)
require(goftest)
require(e1071)

# načítanie dát
dataZ <- read.table("http://www.iam.fmph.uniba.sk/ospm/data/04_data.txt",
header = TRUE)
options(digits=10, scipen=999)
head(dataZ)

Z1 <- dataZ$Z1
Z2 <- dataZ$Z2
Z3 <- dataZ$Z3
Z4 <- dataZ$Z4

# počty uzavretých poistných zmlúv v jednotlivých skupinách
n1 <- length(Z1[!is.na(Z1)])
n2 <- length(Z2[!is.na(Z2)])
n3 <- length(Z3[!is.na(Z3)])
n4 <- length(Z4[!is.na(Z4)])

# nenulové výšky poistných nárokov pre jednotlivé skupiny
Y1 <- Z1[!is.na(Z1) & Z1!=0]
Y2 <- Z2[!is.na(Z2) & Z2!=0]
Y3 <- Z3[!is.na(Z3) & Z3!=0]
Y4 <- Z4[!is.na(Z4) & Z4!=0]

# MODELOVANIE POČTU POISTNÝCH NÁROKOV VYPLÝVAJÚCICH Z JEDNEJ POISTNEJ
# ZMLUVY

# odhadovanie parametrov Bernoulliho (alternatívneho) rozdelenia
# pravdepodobnosti
p1 <- length(Y1)/n1
p2 <- length(Y2)/n2
p3 <- length(Y3)/n3
p4 <- length(Y4)/n4
```

```
# MODELOVANIE NENULOVEJ VÝŠKY POISTNÝCH NÁROKOV

# V tomto programovom kóde uvádzame postup tvorby aktuárskeho modelu
# len pre jednu voľbu rozdelenia, pričom volíme tie rozdelenia
# pravdepodobnosti, ktoré v daných skupinách najpresnejšie opisujú
# správanie sa nenulových výšok poistných nárokov. Postup pri voľbe
# iného rozdelenia pravdepodobnosti je totiž analogický.

# 1. poistná skupina

# základné opisné charakteristiky nenulovej výšky poistných nárokov
mean(Y1)
var(Y1)
sd(Y1)
skewness(Y1)
kurtosis(Y1)
summary(Y1)

# voľba konkrétneho modelu - GAMA ROZDELENIE PRAVDEPODOBNOSTI

# kalibrácia modelu
fit.gamma.Y1 <- fitdist(Y1, distr="gamma",
                        method="mge", discrete=FALSE, gof="CvM")
summary(fit.gamma.Y1)

# validácia modelu
plot(fit.gamma.Y1)
gofstat(fit.gamma.Y1)

ks.test(Y1, "pgamma",
        shape=fit.gamma.Y1$estimate[1],
        rate=fit.gamma.Y1$estimate[2])
cvm.test(Y1, "pgamma",
         shape=fit.gamma.Y1$estimate[1],
         rate=fit.gamma.Y1$estimate[2])
ad.test(Y1, "pgamma",
        shape=fit.gamma.Y1$estimate[1],
        rate=fit.gamma.Y1$estimate[2])

# 2. poistná skupina
# voľba konkrétneho modelu - WEIBULLOVO ROZDELENIE PRAVDEPODOBNOSTI

# kalibrácia modelu
fit.weibull.Y2 <- fitdist(Y2, distr="weibull",
                          method="mge", discrete=FALSE, gof="CvM")
summary(fit.weibull.Y2)

# validacia modelu
plot(fit.weibull.Y2)
gofstat(fit.weibull.Y2)
```

```
#
ks.test(Y2, "pweibull",
        shape=fit.weibull.Y2$estimate[1],
        scale=fit.weibull.Y2$estimate[2])
cvm.test(Y2, "pweibull",
         shape=fit.weibull.Y2$estimate[1],
         scale=fit.weibull.Y2$estimate[2])
ad.test(Y2, "pweibull",
        shape=fit.weibull.Y2$estimate[1],
        scale=fit.weibull.Y2$estimate[2])

# 3. poistná skupina
# voľba konkrétneho modelu - GAMA ROZDELENIE PRAVDEPODOBNOSTI

# kalibrácia modelu
fit.gamma.Y3 <- fitdist(Y3, distr="gamma",
                        method="mge", discrete=FALSE, gof="CvM")
summary(fit.gamma.Y3)

# validácia modelu
plot(fit.gamma.Y3)
gofstat(fit.gamma.Y3)

ks.test(Y3, "pgamma",
        shape=fit.gamma.Y3$estimate[1],
        rate=fit.gamma.Y3$estimate[2])
cvm.test(Y3, "pgamma",
         shape=fit.gamma.Y3$estimate[1],
         rate=fit.gamma.Y3$estimate[2])
ad.test(Y3, "pgamma",
        shape=fit.gamma.Y3$estimate[1],
        rate=fit.gamma.Y3$estimate[2])

# 4. poistná skupina
# voľba konkrétneho modelu - LOG-NORMÁLNE ROZDELENIE PRAVDEPODOBNOSTI

# kalibrácia modelu
fit.lnorm.Y4 <- fitdist(Y4, distr="lnorm",
                        method="mge", discrete=FALSE, gof="CvM")
summary(fit.lnorm.Y4)

# validácia modelu
plot(fit.lnorm.Y4)
gofstat(fit.lnorm.Y4)

ks.test(Y4, "plnorm",
        meanlog=fit.lnorm.Y4$estimate[1],
        sdlog=fit.lnorm.Y4$estimate[2])
cvm.test(Y4, "plnorm",
         meanlog=fit.lnorm.Y4$estimate[1],
         sdlog=fit.lnorm.Y4$estimate[2])
```

```
#
ad.test(Y4, "plnorm",
        meanlog=fit.lnorm.Y4$estimate[1],
        sdlog=fit.lnorm.Y4$estimate[2])

# MONTE CARLO SIMULÁCIA
# nastavenie parametrov
M <- 15000
S.sim <- numeric(M)

set.seed(20221124)
for(t in 1:M) {
  N1.sim <- rbinom(n=n1, size=1, prob=p1)
  Y1.sim <- rgamma(n=n1, shape=fit.gamma.Y1$estimate[1],
                   rate=fit.gamma.Y1$estimate[2])
  Z1.sim <- N1.sim*Y1.sim
  N2.sim <- rbinom(n=n2, size=1, prob=p2)
  Y2.sim <- rweibull(n=n2, shape=fit.weibull.Y2$estimate[1],
                     scale=fit.weibull.Y2$estimate[2])
  Z2.sim <- N2.sim*Y2.sim
  N3.sim <- rbinom(n=n3, size=1, prob=p3)
  Y3.sim <- rgamma(n=n3, shape=fit.gamma.Y3$estimate[1],
                   rate=fit.gamma.Y3$estimate[2])
  Z3.sim <- N3.sim*Y3.sim
  N4.sim <- rbinom(n=n4, size=1, prob=p4)
  Y4.sim <- rlnorm(n=n4, meanlog=fit.lnorm.Y4$estimate[1],
                   sdlog=fit.lnorm.Y4$estimate[2])
  Z4.sim <- N4.sim*Y4.sim
  S.sim[t] <- sum(Z1.sim) + sum(Z2.sim) + sum(Z3.sim) + sum(Z4.sim)
}

# základné opisné charakteristiky simulovanej celkovej výšky poistných
# nárokov
mean(S.sim)
var(S.sim)
sd(S.sim)
skewness(S.sim)
kurtosis(S.sim)

# simulovaná empirická kumulatívna distribučná funkcia
F_S <- ecdf(S.sim)

# NORMÁLNA APROXIMÁCIA
# hodnoty mu a sigma predstavujú strednú hodnotu a disperziu celkovej
# výšky poistných nárokov, ktoré sme vypočítali pomocou odvodených
# vzťahov (4.12), (4.16), (4.17) a (4.18), pričom kód k výpočtu
# v prílohe neuvádzame
mu <- 138721.9328
sigma <- 8814.6559
```

```
# GRAFICKÉ POROVNANIE
# porovnanie simulovanej funkcie hustoty a normálnej aproximácie funkcie
# hustoty (bez znázornenia kvantilov)
curve(dnorm(x, mean=mu, sd=sigma), from=(mu-5*sigma), to=(mu+5*sigma),
      xlab=expression(italic(s)), ylab=expression(italic(f[S](s))),
      col="indianred1", lwd=2, cex.lab=0.8, cex.axis=0.8)
lines(density(S.sim), type="l", col="darkslategrey", lwd=2, lty="longdash")
legend("topleft", c("simulovaná funkcia hustoty", "normálna aproximácia"),
      col=c("darkslategrey", "indianred1"), lty=c("longdash", "solid"),
      lwd=c(2,2), bty="n", cex=0.8)

# 99,5-percentný simulovaný kvantil
q.s <- quantile(F_S, probs=0.995)

# 99,5-percentný kvantil normálnej aproximácie
q.n <- qnorm(0.995, mean=mu, sd=sigma)
```

Príloha B.

Programový kód k Príkladu 5.1 vytvorený v prostredí štatistického softvéru R.

```
# K spusteniu tohto programového kódu je potrebné mať funkčné
# pripojenie k internetu.

# inštalácia potrebných balíkov
install.packages(c("fitdistrplus", "goftest", "e1071"))

# načítanie balíkov
require(fitdistrplus)
require(goftest)
require(e1071)

# načítanie dát
data <- read.table("http://www.iam.fmph.uniba.sk/ospm/data/05_data.txt",
header = TRUE)
options(digits = 10, scipen = 999)
head(data)

# vytvorenie náhodnej premennej W, reprezentujúcej dobu uplynutú
# medzi nahlásením dvoch po sebe nasledujúcich poistných nárokov
T <- data$Ti
W <- rep(0, 600)
W[1] <- T[1]

for (i in 2:600){
  W[i] <- T[i] - T[i-1]
  }

# MODELOVANIE DOBY UPLYNUTEJ MEDZI NAHLÁSENÍM DVOCH PO SEBE
# NASLEDUJÚCICH POISTNÝCH NÁROKOV

# kalibrácia modelu
fit.exp.W <- fitdist(W, distr="exp",
                     method="mge", discrete=FALSE, gof="CvM")
summary(fit.exp.W)

# validácia modelu
plot(fit.exp.W)
gofstat(fit.exp.W)

ks.test(W, "pexp", rate=fit.exp.W$estimate)
cvm.test(W, "pexp", rate=fit.exp.W$estimate)
ad.test(W, "pexp", rate=fit.exp.W$estimate)

# Analogický postup kalibrácie a validácie ostatných rozdelení
# pravdepodobnosti, ktorých vhodnosť nebola po porovnaní
# s exponenciálnym rozdelením dostatočná, v tomto programovom
# kóde neuvádzame.
```

```
# MODELOVANIE VÝŠKY NAHLÁSENÝCH POISTNÝCH NÁROKOV
X <- data$Xi

# základné opisné charakteristiky výšky nahlásených poistných nárokov
mean(X)
var(X)
sd(X)
skewness(X)
kurtosis(X)
summary(X)

# kalibrácia modelu
fit.gamma.X <- fitdist(X, distr="gamma",
                       method="mge", discrete=FALSE, gof="CvM")
summary(fit.gamma.X)

# validácia modelu
plot(fit.gamma.X)
gofstat(fit.gamma.X)

ks.test(X, "pgamma",
        shape=fit.gamma.X$estimate[1],
        rate=fit.gamma.X$estimate[2])
cvm.test(X, "pgamma",
         shape=fit.gamma.X$estimate[1],
         rate=fit.gamma.X$estimate[2])
ad.test(X, "pgamma",
        shape=fit.gamma.X$estimate[1],
        rate=fit.gamma.X$estimate[2])

# Analogický postup kalibrácie a validácie rozdelení pravdepodobnosti,
# ktoré sa javili byť menej vhodné v tomto programovom kóde opäť
# neuvádzame.

# MODELOVANIE HODNOTY PREBYTKU POISŤOVNE A PRAVDEPODOBNOSTI
# ZRUINOVANIA POISŤOVNE

# nastavenie odhadnutých parametrov
lambda <- fit.exp.W$estimate
shape <- fit.gamma.X$estimate[1]
rate <- fit.gamma.X$estimate[2]
EX <- shape/rate

# parametre dané v zadaní Príkladu 5.1
n <- 600
u <- 10000
theta <- 0.20
c <- (1 + theta)*lambda*EX

# počet simulácií
sim.N <- 5      # 1000
sim.P <- 1      # 1000
```

```
#
P.ruin <- numeric(sim.P)+Inf

# nastavenie vykresľovacieho poľa pre vývoj hodnoty prebytku poisťovne
plot(0,0, type="l", xlim=c(0, 2000), ylim=c(-50000, 200000),
     xlab = expression(italic(t)), ylab = expression(italic(U(t))),
     col="white", cex.lab=0.8, cex.axis=0.8)
abline(h=0, lty=2, col="black")

# simulácia náhodného procesu hodnoty prebytku poisťovne
set.seed(2)
for (j in 1:sim.P){
  N.ruin <- numeric(sim.N)+Inf
  T.ruin <- numeric(sim.N)+Inf
  for (k in 1:sim.N){
    W <- rexp(n, lambda)
    T <- cumsum(W)
    X <- rgamma(n, shape, rate)
    S <- cumsum(X)
    U <- u + c*T - S
    U <- c(u, U[1:end(U)[1]-1])
    ruin <- !all(U>=0)
    if (ruin) {
      N.ruin[k] <- min(which(U<0))
      T.ruin[k] <- T[min(which(U<0))]
    }
    # grafické zobrazenie hodnoty prebytku poisťovne
    T.dup <- c(0, rep(T, each=2))
    prijmy <- u + c*T.dup
    vydavky <- rep(c(0, S), each=2)
    spolu <- prijmy - vydavky[1:length(T.dup)]
    if (ruin){
      spolu.pred.ruin <- spolu[1:(N.ruin[k]*2-1)]
      T.dup.pred.ruin <- T.dup[1:(N.ruin[k]*2-1)]
      spolu.po.ruin <- spolu[(N.ruin[k]*2-1):end(spolu)[1]]
      T.dup.po.ruin <- T.dup[(N.ruin[k]*2-1):end(spolu)[1]]
      lines(T.dup.pred.ruin, spolu.pred.ruin, type="l", col="black")
      lines(T.dup.po.ruin, spolu.po.ruin, type="l", col="indianred3")
      points(T.dup[N.ruin[k]*2-1], 0, col="indianred3", pch=19)
    }
    else{
      lines(T.dup, spolu, type="l", xlim=c(0, max(T)),
            ylim=c(-50000, 200000), xlab=expression(italic(t)),
            ylab=expression(italic(U(t))), cex.lab=0.8, col="black")
    }
  }
  N.ruin <- N.ruin[N.ruin<Inf]
  T.ruin <- T.ruin[T.ruin<Inf]
  P.ruin[j] <- length(N.ruin)/sim.N
}
```

```
# číslo nároku, po ktorom priemerne dôjde k zruinovaniu poisťovne
mean(N.ruin)

# priemerný čas zruinovania poisťovne
mean(T.ruin)

# priemerná pravdepodobnosť zruinovania poisťovne
mean(P.ruin)

# Výstupom simulácie je graf, ktorý zobrazuje náhodný proces hodnoty
# prebytku poisťovne. Odporúčame nastaviť parameter sim.N = 5
# a sim.P = 1 z dôvodu prehľadnosti grafu a nízkemu výpočtovému
# zaťaženiu. Pre numerický výstup hodnoty čísla nároku, po ktorom
# priemerne dôjde k zruinovaniu poisťovne, priemerného času
# zruinovania poisťovne a priemernej pravdepodobnosti zruinovania
# poisťovne odporúčame pre replikáciu výsledkov Príkladu 5.1 zvoliť
# nastavenie parametrov sim.N = 1000 a sim.P = 1000.
```