MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Eva Králová

Modelování frekvence nenahlášených škod na úrovni pojistné smlouvy

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Lucie Mazurová, Ph.D.
Studijní program: Finanční matematika

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne ....

Podpis autora

Ráda bych poděkovala vedoucí mé bakalářské práce RNDr. Lucii Mazurové, Ph.D. za cenné rady, ochotu a čas, který mi v průběhu zpracování práce věnovala.

Název práce: Modelování frekvence nenahlášených škod na úrovni pojistné smlouvy

Autor: Eva Králová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Lucie Mazurová, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato práce se zabývá modely pro počet nastalých dosud nenahlášených škod na úrovni pojistné smlouvy. Uvažujeme, že celkový počet škod na smlouvě se řídí Poissonovým nebo negativně binomickým rozdělením. Parametry těchto rozdělení jsou závislé na rizikové expozici, která je v práci zavedena, rovněž popisujeme možné metody jejího stanovení. Odvozujeme rozdělení pro počet nahlášených a nenahlášených škod, obě jsou závislá na době zpoždění nahlášení škody. Pro odhad parametrů těchto rozdělení používáme metodu maximální věrohodnosti. Demonstrujeme fungování modelů prostřednictvím simulační studie.

Klíčová slova: počet nenahlášených škod, rozdělení doby zpoždění, riziková expozice

Title: Modeling the frequency of unreported claims at the policy level

Author: Eva Králová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Lucie Mazurová, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In this thesis we study policy-level models for unreported claim counts. We assume that the total number of claims on a policy follows a Poisson or negative binomial distribution. The parameters of these distributions depend on the risk exposure introduced in the thesis, we also describe possible methods of calculating the risk exposure. We derive distributions for the number of reported and unreported claims, both of which are dependent on the report lag time of a claim. To estimate the parameters of these distributions, we use the maximum likelihood method. We demonstrate the performance of the models via a simulation study.

Keywords: unreported claim count, lag distributions, frequency exposure

Obsah

Úvod

Cílem této bakalářské práce je formulovat pravděpodobnostní modely pro počet nastalých dosud nenahlášených škod a to na úrovni pojistné smlouvy. Pro ni předpokládáme, že celkový počet škod se řídí Poissonovým nebo negativně binomickým rozdělením. V modelech zohledníme závislost na rizikové expozici smlouvy a na rozdělení doby zpoždění. Modely ilustrujeme na simulační studii.

Práce se skládá ze čtyř kapitol, první tři tvoří teoretickou část, čtvrtá se věnuje simulační studii. Primárním zdrojem pro teoretickou část je Landry a Martin (2022).

V první kapitole zavedeme potřebná rozdělení a odvodíme rozdělení pro počet nahlášených škod. Popíšeme, jakým způsobem lze odhadnout parametry tohoto rozdělení z údajů o již nahlášených škodách, za předpokladu, že známe pravděpodobnost, že škoda byla nahlášena do data vyhodnocení. Představíme v ní také pojem riziková expozice a uvedeme možné metody jejího stanovení. Pozměníme způsob pro odhad parametrů rozdělení nahlášených škod zohledněním rizikové expozice.

Ve druhé kapitole pojednáme o způsobu určení pravděpodobnosti, že škoda byla nahlášena do data vyhodnocení. Dále definujeme rozdělení doby zpoždění a ukážeme, jakým postupem odhadneme parametry tohoto rozdělení z dat o nahlášených škodách.

Třetí kapitola se věnuje odvození rozdělení počtu nastalých dosud nenahlášených škod.

V kapitole se simulační studií ilustrujeme některé uvažované modely a mezi sebou je porovnáme. Data o pojistných smlouvách a škodách simulujeme pomocí softwaru *Wolfram Mathematica* (verze 13.1 Student Edition).

1. Rozdělení počtu nahlášených škod

V této kapitole zavedeme potřebné charakteristiky pojistné smlouvy, odvodíme rozdělení počtu nahlášených škod a ukážeme, jak lze odhadnout parametry tohoto rozdělení. Představíme pojem riziková expozice a ukážeme příklady jejího stanovení.

1.1 Model s jednou pojistnou smlouvou

Uvažujme čas vyhodnocení T a pojistnou smlouvu A takovou, že konec její platnosti nastal před T. Pro smlouvu A označíme jako N celkový počet škod, jako X počet škod, které byly do času T nahlášeny a jako q pravděpodobnost, že škoda na smlouvě A byla nahlášena do času T. Předpokládáme, že škody na smlouvě A nastávají navzájem nezávisle.

Dále předpokládáme, že náhodná veličina N má Poissonovo rozdělení s parametrem $\lambda > 0$ s pravděpodobnostní funkcí

$$P[N = i] = \frac{\lambda^i}{i!} e^{-\lambda}, \quad i = 0, 1, 2, \ldots, \tag{1.1}$$

nebo negativně binomické rozdělení s parametry $k \in (0,\infty)$ a $p \in (0,1)$ s pravděpodobnostní funkcí

$$P[N = i] = \binom{k+i-1}{i} p^k (1-p)^i, \quad i = 0, 1, 2, \ldots, \tag{1.2}$$

přičemž v tomto případě definujeme kombinační číslo jako

$$\binom{m}{r} = \frac{m(m-1)\cdots(m-r+1)}{r!}, \quad r \in \mathbb{N}, m \in \mathbb{R}.$$

Použitím vztahu $\Gamma(m+1) = m\Gamma(m)$, $m \in \mathbb{R}$, a $(r-1)! = \Gamma(r)$, $r \in \mathbb{N}$ můžeme kombinační číslo $\binom{m}{r}$ vyjádřit jako

$$\frac{m(m-1)\cdots(m-r+1)}{r!} \frac{\Gamma(m-r+1)}{\Gamma(m-r+1)} = \frac{\Gamma(m+1)}{\Gamma(r+1)\Gamma(m-r+1)}. \tag{1.3}$$

Střední hodnota náhodné veličiny N s Poissonovým rozdělením s parametrem λ je

$$\mathsf{E}N = \lambda, \tag{1.4}$$

pro náhodnou veličinu N s negativně binomickým rozdělením s parametry k a p má střední hodnota tvar

$$\mathsf{E}N = \frac{k(1-p)}{p}. \tag{1.5}$$

Podmíněné rozdělení náhodné veličiny X za podmínky $N = n$ je binomické s parametry n a q. Binomické rozdělení s parametry $n \in \mathbb{N}$ a $q \in (0,1)$ má pravděpodobnostní funkci

$$P[X = i] = \binom{n}{i} q^i (1-q)^{n-i}, \quad i = 0, 1, 2, \ldots. \tag{1.6}$$

Všechny výše uvedené pravděpodobnostní funkce a střední hodnoty náhodných veličin jsme převzali z DeGroot a Schervish (2012).

Pravděpodobnost q nyní považujeme za známou, o jejím odhadu z údajů o nahlášených škodách pojednáme v kapitole 2. Uvedeme dvě tvrzení popisující vztah mezi N a X, které nám umožní zjistit parametry rozdělení N za předpokladu, že rozdělení X známe. Tvrzení vycházejí z odvození uvedených v Landry a Martin (2022).

Tvrzení 1. *Nechť má náhodná veličina N Poissonovo rozdělení s parametrem $\lambda > 0$ a podmíněné rozdělení náhodné veličiny X za podmínky $N = n$ je binomické rozdělení s parametry $n \in \mathbb{N}$ a $q \in (0,1)$, pak nepodmíněné rozdělení náhodné veličiny X je Poissonovo rozdělení s parametrem $q\lambda$.*

Důkaz. Využijeme větu o úplné pravděpodobnosti a skutečnost, že pro $i \in \mathbb{N}_0$, $j \in \mathbb{N}_0$ takové, že $i < j$ platí $P(X = j \mid N = i) = 0$, tedy

$$P(X = j) = \sum_{i=0}^{\infty} P(N = i)P(X = j \mid N = i) = \sum_{i=j}^{\infty} P(N = i)P(X = j \mid N = i). \tag{1.7}$$

Při použití zavedených pravděpodobnostních funkcí (1.1), (1.6) a definice kombinačního čísla dostáváme

$$\begin{aligned} P(X = j) &= \sum_{i=j}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \binom{i}{j} q^j (1-q)^{i-j} \\ &= \sum_{i=j}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{i!}{(i-j)!j!} q^j (1-q)^{i-j} = \frac{q^j e^{-\lambda}}{j!} \sum_{i=j}^{\infty} \frac{\lambda^i (1-q)^{i-j}}{(i-j)!}. \end{aligned}$$

Označíme-li $a = i - j$, pak můžeme psát

$$P(X = j) = \frac{q^j e^{-\lambda}}{j!} \sum_{a=0}^{\infty} \frac{\lambda^{a+j}(1-q)^a}{a!} = \frac{(q\lambda)^j e^{-\lambda}}{j!} \sum_{a=0}^{\infty} \frac{(\lambda(1-q))^a}{a!}.$$

Po vynásobení výrazem $e^{\lambda(1-q)}e^{-\lambda(1-q)}$ dostáváme

$$\begin{aligned} P(X = j) &= e^{\lambda(1-q)} \frac{(q\lambda)^j e^{-\lambda}}{j!} \sum_{a=0}^{\infty} \frac{(\lambda(1-q))^a}{a!} e^{-\lambda(1-q)} \\ &= \frac{(q\lambda)^j e^{-q\lambda}}{j!} \sum_{a=0}^{\infty} \frac{(\lambda(1-q))^a}{a!} e^{-\lambda(1-q)} = \frac{(q\lambda)^j e^{-q\lambda}}{j!} \sum_{a=0}^{\infty} P(Y = a) \\ &= \frac{(q\lambda)^j e^{-q\lambda}}{j!}, \end{aligned}$$

kde jako Y označujeme náhodnou veličinu s Poissonovým rozdělením s parametrem $\lambda(1-q)$. Tedy X má Poissonovo rozdělení s parametrem $q\lambda$.

□

Tvrzení 2. *Nechť má náhodná veličina N negativně binomické rozdělení s parametry $k \in (0,\infty)$ a $p \in (0,1)$ a podmíněné rozdělení náhodné veličiny X za podmínky $N = n$ je binomické rozdělení s parametry $n \in \mathbb{N}$ a $q \in (0,1)$, pak nepodmíněné rozdělení náhodné veličiny X je negativně binomické rozdělení s parametry k a $\frac{p}{p+q-pq}$.*

Důkaz. Stejným způsobem jako v důkaze tvrzení 1 dostaneme (1.7), použijeme zavedené pravděpodobností funkce (1.2), (1.6) a vztah (1.3), takže dostáváme

$$P(X=j)=\sum_{i=j}^{\infty}\frac{\Gamma(i+k)}{\Gamma(i+1)\Gamma(k)}p^k(1-p)^i\binom{i}{j}q^j(1-q)^{i-j}.$$

Z definice kombinačního čísla a vztahu $(b-1)!=\Gamma(b)$, $b\in\mathbb{N}$ dále plyne

$$P(X=j)=\sum_{i=j}^{\infty}\frac{\Gamma(i+k)}{\Gamma(i+1)\Gamma(k)}p^k(1-p)^i\frac{\Gamma(i+1)}{\Gamma(j+1)\Gamma(i-j+1)}q^j(1-q)^{i-j}$$
$$=\frac{p^kq^j}{\Gamma(k)\Gamma(j+1)}\sum_{i=j}^{\infty}\frac{\Gamma(i+k)}{\Gamma(i-j+1)}(1-p)^i(1-q)^{i-j}.$$

Označíme-li $a=i-j$, pak můžeme psát

$$P(X=j)=\frac{p^kq^j}{\Gamma(k)\Gamma(j+1)}\sum_{a=0}^{\infty}\frac{\Gamma(a+j+k)}{\Gamma(a+1)}(1-p)^{a+j}(1-q)^a. \tag{1.8}$$

Po vynásobení výrazem $\Gamma(j+k)/\Gamma(j+k)$ máme

$$P(X=j)=\frac{p^kq^j(1-p)^j}{\Gamma(k)\Gamma(j+1)}\sum_{a=0}^{\infty}\frac{\Gamma(a+j+k)}{\Gamma(a+1)}((1-p)(1-q))^a\frac{\Gamma(j+k)}{\Gamma(j+k)}$$
$$=\frac{\Gamma(j+k)}{\Gamma(k)\Gamma(j+1)}p^kq^j(1-p)^j\sum_{a=0}^{\infty}\frac{\Gamma(a+j+k)}{\Gamma(a+1)\Gamma(j+k)}((1-p)(1-q))^a.$$

Vynásobením výrazem $(1-(1-p)(1-q))^{j+k}/(1-(1-p)(1-q))^{j+k}$ dostáváme

$$P(X=j)=\frac{\Gamma(j+k)}{\Gamma(k)\Gamma(j+1)}\frac{p^kq^j(1-p)^j}{(1-(1-p)(1-q))^{j+k}}$$
$$\times\sum_{a=0}^{\infty}\frac{\Gamma(a+j+k)}{\Gamma(a+1)\Gamma(j+k)}((1-p)(1-q))^a(1-(1-p)(1-q))^{j+k}.$$

Přitom platí

$$\frac{p^kq^j(1-p)^j}{(1-(1-p)(1-q))^{j+k}}=\frac{p^kq^j(1-p)^j}{(p+q-pq)^{j+k}}=\left(\frac{p}{p+q-pq}\right)^k\left(\frac{q(1-p)}{p+q-pq}\right)^j,$$

tedy můžeme (1.8) vyjádřit ve tvaru

$$P(X=j)=\frac{\Gamma(j+k)}{\Gamma(k)\Gamma(j+1)}\left(\frac{p}{p+q-pq}\right)^k\left(\frac{q(1-p)}{p+q-pq}\right)^j\sum_{a=0}^{\infty}P(Y=a)$$
$$=\frac{\Gamma(j+k)}{\Gamma(k)\Gamma(j+1)}\left(\frac{p}{p+q-pq}\right)^k\left(\frac{q(1-p)}{p+q-pq}\right)^j,$$

kde jako Y označujeme náhodnou veličinu s negativně binomickým rozdělením s parametry $j+k$ a $(1-p)(1-q)$. Tedy X má negativně binomické rozdělení s parametry k a $\frac{p}{p+q-pq}$.

□

1.2 Model s vícero pojistnými smlouvami

Nyní uvažujme skupinu nezávislých pojistných smluv $A_1, A_2, \dots, A_m$, $m \in \mathbb{N}$, ve které u každé smlouvy A_i, $i = 1, 2, \dots, m$ nastal konec platnosti před časem T. Předpokládáme, že škody na smlouvě A_i nastávají navzájem nezávisle. Označíme jako N_i celkový počet škod na pojistné smlouvě A_i, jako X_i počet škod na pojistné smlouvě A_i, které byly nahlášeny do času T a jako q_i pravděpodobnost, že škoda vztahující se k pojistné smlouvě A_i byla nahlášena do času T.

Nejprve předpokládejme, že N_i, $i = 1, 2, \dots, m$ mají všechny Poissonovo rozdělení s parametrem λ. Víme tedy, že podmíněné rozdělení náhodné veličiny X_i za podmínky $N_i = n$ má binomické rozdělení s parametry n a q_i, z tvrzení 1 vyplývá, že nepodmíněné rozdělení náhodné veličiny X_i je Poissonovo rozdělení s parametrem λq_i pro $i = 1, 2, \dots, m$.

Dále předpokládáme, že v čase T máme pozorování $X_1 = x_1, \dots, X_m = x_m$ a pravděpodobnosti q_i považujeme za známé (odhadujeme je v kapitole 2), proto můžeme použít metodu maximální věrohodnosti pro odhad parametru λ. Logaritmická věrohodnostní funkce pro parametr λ je tvaru

$$\begin{aligned}\ell(\lambda) &= \sum_{i=1}^{m} \log(P(X_i = x_i)) = \sum_{i=1}^{m} \log\left(\frac{(\lambda q_i)^{x_i}}{x_i!} e^{-\lambda q_i}\right) \\ &= \sum_{i=1}^{m} (x_i \log(\lambda q_i) - \lambda q_i - \log(x_i!)).\end{aligned}$$

Nyní předpokládejme, že N_i, $i = 1, \dots, m$ mají všechny negativně binomické rozdělení s parametry k a p. Pak z tvrzení 2 vyplývá, že nepodmíněné rozdělení náhodné veličiny X_i je negativně binomické rozdělení s parametry k a $\frac{p}{p+q_i-pq_i}$ pro $i = 1, 2, \dots, m$ a logaritmická věrohodnostní funkce pro odhad parametrů k, p má následující tvar

$$\ell(k, p) = \sum_{i=1}^{m} \log\left(\binom{x_i + k - 1}{x_i} \left(\frac{p}{p + q_i - pq_i}\right)^k \left(1 - \left(\frac{p}{p + q_i - pq_i}\right)\right)^{x_i}\right).$$

1.3 Riziková expozice

Předpoklad, že rozdělení celkového počtu škod má u každé smlouvy stejné parametry, je v praxi nerealistický. Proto je třeba zavést veličinu, která bude reflektovat rozdíly jednotlivých smluv. Takovou veličinu zavedli Landry a Martin (2022) pod názvem „frequency exposure“, my budeme používat označení riziková expozice. Předpokládáme, že střední hodnota celkového počtu škod nastalých na pojistné smlouvě je úměrná rizikové expozici této smlouvy.

Pro každou pojistnou smlouvu A_i, $i = 1, 2, \dots, m$ určíme rizikovou expozici $e_i \in (0,\infty)$ podle způsobu uvedeného v podkapitole 1.4. Ukážeme, jakým způsobem parametry rozdělení celkového počtu škod na smlouvě A_i závisí na rizikové expozici e_i.

Opět začneme s případem, kdy se celkový počet škod řídí Poissonovým rozdělením. Předpokládejme, že střední hodnota celkového počtu škod na smlouvě s jednotkovou rizikovou expozicí je λ. Z předpokladu

$$\mathsf{E} N_i = e_i \lambda = \lambda_i,$$

vyplývá, že počet škod N_i na smlouvě s expozicí e_i má Poissonovo rozdělení s parametrem λe_i. Použitím tvrzení 1 dostáváme, že X_i má Poissonovo rozdělení s parametrem $\lambda e_i q_i$. Modifikovaná logaritmická věrohodnostní funkce pro parametr λ je tvaru

$$\ell(\lambda) = \sum_{i=1}^{m} \ln\left(\frac{(\lambda q_i e_i)^{x_i}}{x_i!} e^{-\lambda q_i e_i}\right) = \sum_{i=1}^{m} (x_i \ln(\lambda q_i e_i) - \lambda q_i e_i - \ln(x_i!)). \tag{1.9}$$

Nechť má nyní celkový počet škod na smlouvě negativně binomické rozdělení. Opět předpokládáme, že střední hodnota celkového počtu škod na smlouvě A_i je úměrná rizikové expozici. Budeme teď předpokládat, že počty škod N_i na jednotlivých smlouvách budou mít negativně binomické rozdělení se stejným parametrem k a s parametry p_i závislým na rizikové expozici. Nechť p je parametr pro rozdělení škod při jednotkové expozici. Střední hodnota negativně binomického rozdělení s parametry k a p je (1.5), tedy předpokládáme

$$\mathsf{E}N_i = e_i \, k\frac{1-p}{p}. \tag{1.10}$$

Zároveň platí

$$\mathsf{E}N_i = k\frac{1-p_i}{p_i},$$

odkud porovnáním s (1.10) dostáváme

$$p_i = \frac{p}{p + e_i - pe_i}.$$

Tedy N_i se řídí negativně binomickým rozdělením s parametry k a $p/(p+e_i-pe_i)$. Z tvrzení 2 vyplývá, že X_i má negativně binomické rozdělení s parametry k a $p_i/(p_i + q_i - p_i q_i)$. Úpravami dostáváme

$$\frac{p_i}{p_i + q_i - p_i q_i} = \frac{\frac{p}{p+e_i-pe_i}}{\frac{p}{p+e_i-pe_i} + q_i - \frac{p}{p+e_i-pe_i} q_i} = \frac{p}{p + q_i(p + e_i - pe_i) - pq_i}.$$

Tedy X_i má negativně binomické rozdělení s parametry k a $p/(p + e_i q_i - pe_i q_i)$. Modifikovaná logaritmická věrohodnostní funkce pro parametry k a p je tvaru

$$\ell(k,p) = \sum_{i=1}^{m} \ln\left(\binom{x_i + k - 1}{x_i}\left(\frac{p}{p + q_i e_i - pq_i e_i}\right)^{k}\left(1 - \left(\frac{p}{p + q_i e_i - pq_i e_i}\right)\right)^{x_i}\right). \tag{1.11}$$

1.4 Metody určení rizikové expozice

Podle Landry a Martin (2022) lze rizikovou expozici stanovit na základě pojistné doby, zaslouženého pojistného nebo výše spoluúčasti.

1. Očekávaný počet škod je přímo úměrný délce působení pojistné smlouvy. Hodnotu e_i můžeme definovat jako $V_i/365$, kde V_i je počet dní, po které je do času T smlouva A_i v platnosti. Tedy nemusíme již uvažovat, že konec platnosti smlouvy nastal před datem T. Zřejmě je $e_i = 1$ pro $V_i = 365$, tedy pojistnou smlouvu, která je v působení jeden rok.

2. Očekávaný počet škod je přímo úměrný zaslouženému pojistnému.
Stanovíme základní pojistné pod označením ZP, které se bude vztahovat k jednotkové rizikové expozici. Nechť P_i je výše zaslouženého pojistného u pojistné smlouvy A_i, pak $e_i = P_i/ZP$.

3. Očekávaný počet škod závisí na výši spoluúčasti.
Můžeme předpokládat, že se zvyšující se spoluúčastí se bude snižovat počet škod, které smlouva kryje, protože se snižuje pravděpodobnost, že výše škody překročí spoluúčast. Označme jako D_i výši spoluúčasti pro pojistnou smlouvu A_i. Uvažujme, že rozdělení výší škod je pro každou smlouvu A_i stejné, označme jako $F(x)$, $x \geq 0$ distribuční funkci tohoto rozdělení, potom $e_i = 1 - F(D_i)$, totiž $1 - F(D_i)$ představuje pravděpodobnost, že výše škody překročí spoluúčast V tomto případě dostáváme jednotkovou rizikovou expozici, když je spoluúčast nulová $(1 - F(0) = 1)$.

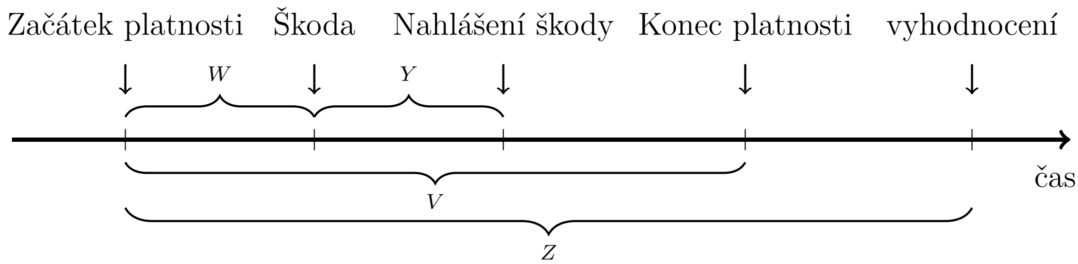
2. Odhad pravděpodobnosti nahlášení škody

V této kapitole ukážeme, jak lze odhadnout q_i, tedy pravděpodobnost, že škoda na smlouvě A_i bude nahlášena do času vyhodnocení T.

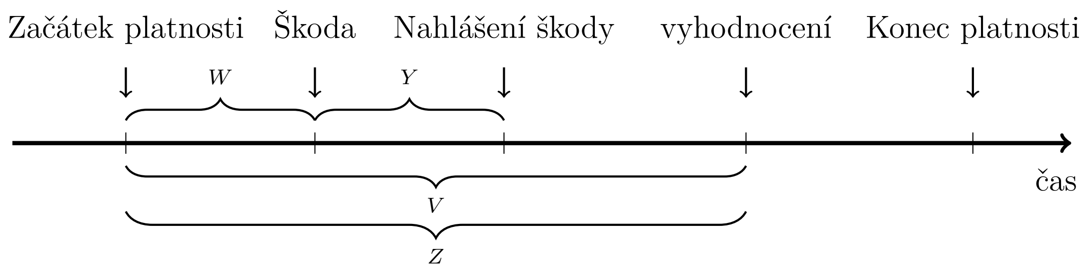
2.1 Pravděpodobnost q_i

Dobu zpoždění Y definujeme jako rozdíl data, kdy byla škoda nahlášena, a data, kdy škoda nastala. Předpokládejme, že Y je náhodná veličina se spojitým rozdělením s distribuční funkcí $R(x)$, $x \geq 0$, a hustotou $r(x)$, $x \geq 0$.

Pro pojistnou smlouvu A označíme jako V rozdíl data vyhodnocení T a data začátku pojištění v případě, kdy nastane datum vyhodnocení dříve než datum konce pojištění, nebo jako rozdíl data konce pojištění a data začátku pojištění v případě opačném. Označme jako Z dobu od data začátku pojistné smlouvy A po datum vyhodnocení. Pro škodu nastalou na pojistné smlouvě A označíme jako W dobu mezi datem začátku pojištění a datem nastání škody. Graficky znázorníme tyto hodnoty na časové ose pro dva případy:



Obrázek 2.1: Průběh pojištění v případě, kdy datum vyhodnocení následuje po datu konce pojištění.



Obrázek 2.2: Průběh pojištění v případě, kdy konec pojištění následuje po datu vyhodnocení.

Dále budeme předpokládat, že škody nastávají rovnoměrně během pojistného období, tedy W má rovnoměrné rozdělení na intervalu $(0,V)$. Pak pravděpodobnost, že škoda bude nahlášena do data vyhodnocení můžeme vyjádřit jako $P(W+Y \leq Z)$. Z věty 8.1 ze Zvára (1997)(str. 116) plyne

$$P(W+Y \leq Z) = \int_0^V R(Z-w)\frac{1}{V}dw = \int_0^V \int_0^{Z-w} r(y)\frac{1}{V}dydw. \tag{2.1}$$

Tento dvojný integrál pro některá rozdělení doby zpoždění nelze přesně vyjádřit (například pro gama rozdělení) a pro velký počet smluv by kalkulace byla výpočetně náročná, proto zavedeme další předpoklad, který situaci zjednoduší.

Pokud budeme předpokládat, že všechny škody nastaly v polovině doby trvání pojistné smlouvy, W již nebude náhodná veličina a bude platit $W = \frac{V}{2}$. Potom

$$\begin{aligned} P(W+Y \leq Z) &= P(Y \leq Z - W) \\ &= P(Y \leq Z - V/2) \\ &= R(Z - V/2). \end{aligned}$$

Výše zavedené symboly V a Z se nyní budou vztahovat ke smlouvě A_i, což vyznačíme indexem i. Potom

$$q_i = R\left(Z_i - \frac{V_i}{2}\right). \tag{2.2}$$

2.2 Rozdělení doby zpoždění

Vhodnou volbou pro rozdělení doby zpoždění mohou být exponenciální nebo gama rozdělení. Hustota gama rozdělení s parametry $a > 0$ a $p > 0$ je tvaru

$$f(x) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}, \quad x \geq 0.$$

Exponenciální rozdělení s parametrem $\mu > 0$ má hustotu ve tvaru

$$f(x) = \mu e^{-\mu x}, \quad x \geq 0. \tag{2.3}$$

Obě hustoty jsme převzali z DeGroot a Schervish (2012).

Mějme škody S_i nahlášené do data vyhodnocení T, $i = 1, 2, \ldots, l$, kde l odpovídá počtu všech nahlášených škod. Abychom mohli odhadnout parametry rozdělení doby zpoždění, budeme potřebovat o těchto škodách následující údaje: dobu zpoždění nahlášení Y_i škody S_i a rozdíl data vyhodnocení T a jejího data nastání, ten označíme G_i. Mějme pozorování $Y_1 = y_1, \ldots, Y_l = y_l$ a $G_1 = g_1, \ldots, G_l = g_l$. Jelikož máme k dispozici informace o škodách jen do data vyhodnocení T, jsou naše údaje zprava useknuté, to tedy musí platit i pro rozdělení doby zpoždění. Tuto vlastnost je potřeba při odhadu parametrů zohlednit. Musíme tedy uvažovat podmíněné rozdělení doby zpoždění za podmínky $x \leq g_i$. Podmíněnou distribuční funkci rozdělení doby zpoždění získáme pro $i = 1, 2, \ldots, l$ jako

$$P(Y_i \leq x \mid Y_i \leq g_i) = \frac{P(Y_i \leq x \cap Y_i \leq g_i)}{P(Y_i \leq g_i)} = \frac{P(Y_i \leq x)}{P(Y_i \leq g_i)} = \frac{R(x)}{R(g_i)}.$$

Hustotu získáme zderivováním podle x, tedy $\frac{r(x)}{R(g_i)}$. Logaritmická věrohodnostní funkce pro vektor parametrů rozdělení doby zpoždění $\boldsymbol{\theta}$ je tvaru

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{l} \log\left(\frac{r(y_i)}{R(g_i)}\right). \tag{2.4}$$

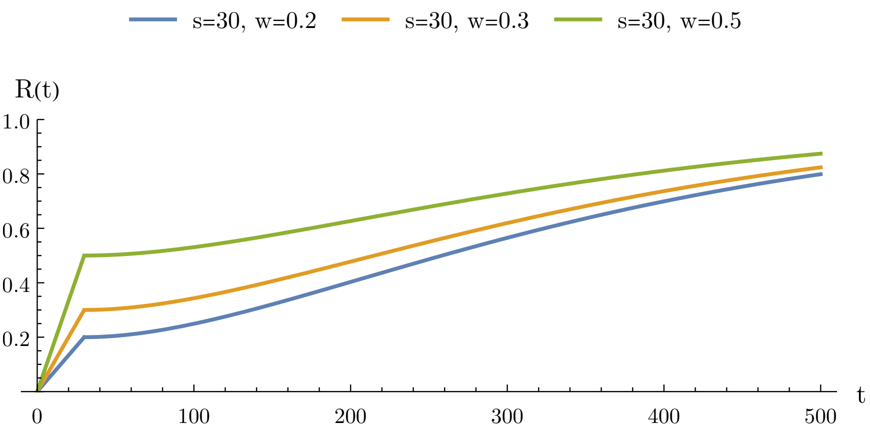
2.3 Kombinované rozdělení doby zpoždění

Rozdělení doby zpoždění by mělo zohlednit, že pro mnoho typů pojištění je vysoká pravděpodobnost, že škoda bude nahlášena relativně rychle po době nastání škody. Zároveň musí zachycovat, že existuje signifikantní pravděpodobnost, že škoda bude nahlášena později. Tyto požadavky splňuje kombinace složená z rovnoměrného rozdělení a posunutého exponenciálního nebo gama rozdělení. Uvedeme distribuční funkci a hustotu kombinace rozdělení:

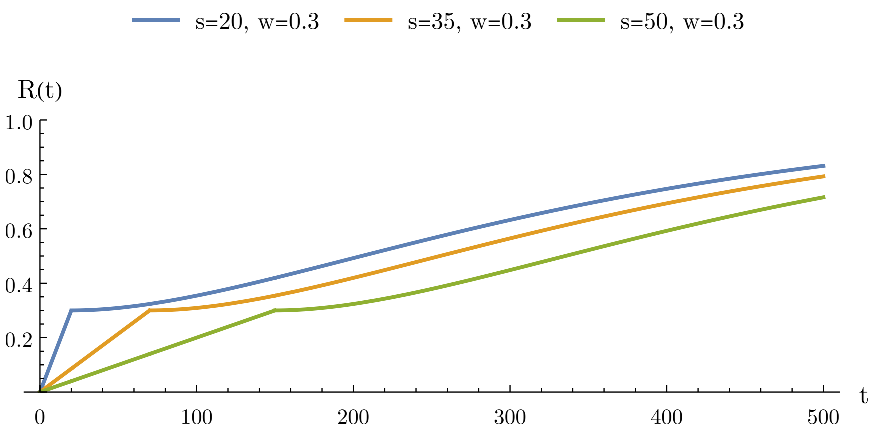
$$R(t) = \begin{cases} 0, & t \leq 0, \\ w\dfrac{t}{s}, & 0 < t \leq s, \\ w + (1-w)H(t-s), & t > s, \end{cases}$$

$$r(t) = \begin{cases} 0, & t \leq 0, \\ w\dfrac{1}{s}, & 0 < t \leq s, \\ w + (1-w)h(t-s), & t > s, \end{cases}$$

kde $h(x)$ označuje hustotu a $H(x)$ distribuční funkci exponenciálního nebo gama rozdělení. Váhu rovnoměrného rozdělení určuje w a s je pravý okraj nosiče rovnoměrného rozdělení. Na obrázku 2.3 ilustrujeme vliv parametru w a na obrázku 2.4 vliv parametru s na tvar distribuční funkce doby zpoždění. Dobu zpoždění uvažujeme ve dnech.



Obrázek 2.3: Distribuční funkce kombinace rovnoměrného rozdělení na intervalu (0, 30) a gama rozdělení s parametry 2 a 175 pro různé parametry w.



Obrázek 2.4: Distribuční funkce kombinace rovnoměrného rozdělení na intervalu $(0, s)$ a gama rozdělení s parametry 2 a 175 pro různé parametry s.

3. Rozdělení počtu nastalých dosud nenahlášených škod

Cílem této kapitoly je odvodit rozdělení počtu nastalých dosud nenahlášených škod. To nám umožňují dvě následující tvrzení, které vychází z odvození z Landry a Martin (2022).

Tvrzení 3. *Nechť má náhodná veličina N Poissonovo rozdělení s parametrem $\lambda > 0$ a podmíněné rozdělení náhodné veličin X za podmínky $N = n$ je binomické rozdělení s parametry $n \in \mathbb{N}$ a $q \in (0,1)$, pak podmíněné rozdělení $N - X$ za podmínky $X = x$ je Poissonovo rozdělení s parametrem $\lambda(1 - q)$.*

Důkaz. Využijeme definici podmíněné pravděpodobnosti a dostáváme

$$P(N = i \mid X = j) = \frac{P(N = i \cap X = j)}{P(X = j)} = \frac{P(N = i)P(X = j \mid N = i)}{P(X = j)}. \tag{3.1}$$

Použijeme zavedené pravděpodobnostní funkce (1.1), (1.6), tvrzení 1, definici kombinačního čísla a pro $i = j,\, j + 1, \ldots$ přepíšeme (3.1) jako

$$P(N = i \mid X = j) = \frac{\frac{e^{-\lambda}\lambda^i}{i!}\binom{i}{j}q^j(1-q)^{i-j}}{\frac{e^{-q\lambda}(q\lambda)^j}{j!}} = \frac{\frac{e^{-\lambda}\lambda^i}{i!}\frac{i!}{(i-j)!j!}q^j(1-q)^{i-j}}{\frac{e^{-q\lambda}(q\lambda)^j}{j!}}$$

$$= e^{-\lambda+q\lambda}\lambda^{i-j}\frac{1}{(i-j)!}(1-q)^{i-j} = \frac{e^{-(1-q)\lambda}((1-q)\lambda)^{i-j}}{(i-j)!}.$$

Označíme-li $a = i - j$, pak pro $a = 0,\, 1, \ldots$ můžeme psát

$$\begin{aligned} P(N - X = a \mid X = j) &= P(N = a + X \mid X = j) \\ &= \frac{e^{-(1-q)\lambda}((1-q)\lambda)^a}{a!}. \end{aligned}$$

□

Tvrzení 4. *Nechť náhodná veličina N má negativně binomické rozdělení s parametry $k \in (0,\infty)$ a $p \in (0,1)$ a podmíněné rozdělení náhodné veličiny X za podmínky $N = n$ je binomické rozdělení s parametry $n \in \mathbb{N}$ a $q \in (0,1)$, pak podmíněné rozdělení $N - X$ za podmínky $X = x$ je negativně binomické rozdělení s parametry $k + x$ a $p + q - pq$.*

Důkaz. Stejným způsobem jako v důkaze tvrzení 3 dostaneme(3.1), použijeme zavedené pravděpodobnostní funkce (1.2), (1.6), tvrzení 2 a pro $i = j,\, j + 1, \ldots$ nyní dostáváme (3.1) ve tvaru

$$P(N = i \mid X = j) = \frac{\frac{\Gamma(i+k)}{\Gamma(i+1)\Gamma(k)}p^k(1-p)^i\binom{i}{j}q^j(1-q)^{i-j}}{\frac{\Gamma(j+k)}{\Gamma(j+1)\Gamma(k)}\left(\frac{p}{p+q-pq}\right)^k\left(\frac{q-qp}{p+q-pq}\right)^j}. \tag{3.2}$$

Užitím definice kombinačního čísla a vztahu $(b-1)! = \Gamma(b)$, $b \in \mathbb{N}$ dále upravíme: (3.2)

$$\begin{aligned}
&\frac{\Gamma(i+k)}{\Gamma(i+1)\Gamma(k)} \frac{\Gamma(j+1)\Gamma(k)}{\Gamma(j+k)} \frac{\Gamma(i+1)}{\Gamma(j+1)\Gamma(i-j+1)} \frac{p^k(1-p)^i q^j (1-q)^{i-j}}{\frac{p^k q^j (1-p)^j}{(p+q-pq)^{k+j}}} \\
&= \frac{\Gamma(i+k)(1-p)^{i-j}(1-q)^{i-j}}{\Gamma(i-j+1)\Gamma(j+k)} (p+q-pq)^{k+j} \\
&= \frac{\Gamma((i-j)+(k+j))(1-p-q+pq)^{i-j}}{\Gamma(i-j+1)\Gamma(j+k)} (p+q-pq)^{k+j}
\end{aligned}$$

Označíme-li $a = i - j$, pak pro $a = 0, 1, \ldots$ můžeme psát

$$\begin{aligned}
&P(N - X = a \mid X = j) = P(N = a + X \mid X = j) \\
&= \frac{\Gamma(a+(k+j))}{\Gamma(a+1)\Gamma(k+j)} (p+q-pq)^{k+j} (1-p-q+pq)^a.
\end{aligned}$$

□

Mějme opět soubor pojistných smluv A_i, $i = 1, \ldots, m$. V prvním případě, kdy N_i má Poissonovo rozdělení s parametrem λe_i, pak z tvrzení 3 vyplývá, že počet nenahlášených škod má Poissonovo rozdělení s parametrem

$$\lambda e_i (1 - q_i). \tag{3.3}$$

Tedy v případě, kdy předpokládáme, že celkový počet škod se řídí Poissonovým rozdělením, není rozdělení počtu nenahlášených škod závislé na počtu škod do času T nahlášených. Když se N_i řídí negativně binomickým rozdělením s parametry k a $\frac{p}{p+e_i-pe_i}$, pak v případě, že do času vyhodnocení T bylo na smlouvě A_i nahlášeno x_i škod, má počet nenahlášených škod na smlouvě A_i negativně binomické rozdělení s parametry

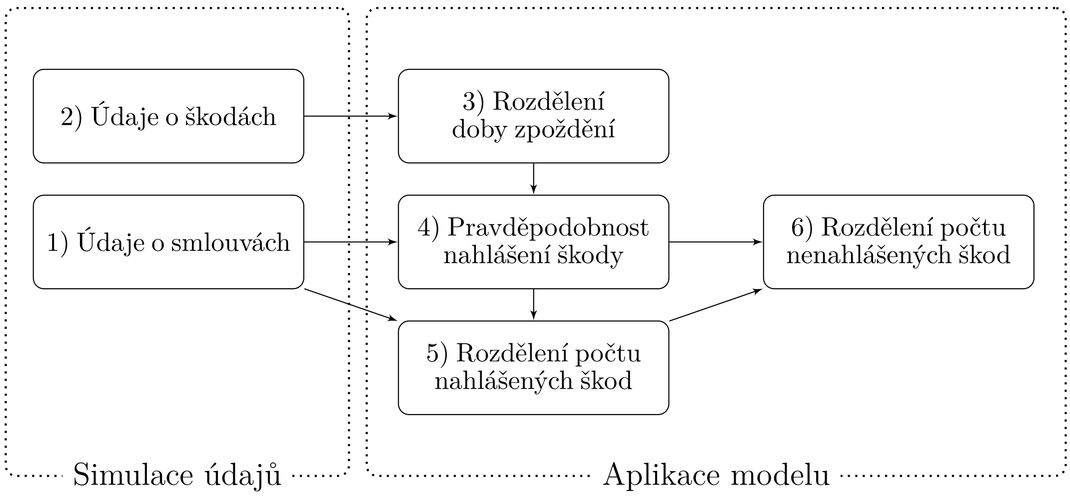
$$k + x_i \text{ a } p_i + q_i - p_i q_i, \tag{3.4}$$

kde $p_i = \frac{p}{p+e_i-pe_i}$.

4. Simulační studie

V této kapitole ilustrujeme praktické využití teorie z prvních tří kapitol pro odhad rozdělení počtu nastalých dosud nenahlášených škod. Na všechny výpočty používáme software Wolfram Mathematica (verze 13.1 Student Edition). Veškeré vstupy a výstupy jsou k nahlédnutí v příloze k práci.

Nejdříve simulujeme údaje o pojistných smlouvách a škodách, které byly nahlášeny do data vyhodnocení T. U pojistných smluv se jedná o data počátku a konce jejich platnosti a kolik škod na nich bylo do data vyhodnocení T nahlášeno. U nahlášených škod simulujeme data nastání a nahlášení. Z údajů o nahlášených škodách budeme odhadovat parametry rozdělení doby zpoždění. Parametry rozdělení pro počet nastalých dosud nenahlášených škod odhadneme ze simulovaných údajů. Následující obrázek znázorňuje návaznost a propojení jednotlivých kroků:



Obrázek 4.1: Proces odhadování rozdělení počtu nastalých dosud nenahlášených škod.

4.1 Simulace údajů o pojistných smlouvách a škodách

Budeme uvažovat skupinu pojistných smluv $A_1, A_2, \ldots, A_{1000}$. Celkový počet škod na smlouvě s jednotkovou rizikovou expozicí budeme generovat z negativně binomického rozdělení s parametry $\tilde{k} = 0.5$ a $\tilde{p} = 0.5$. Dobu zpoždění nahlášení každé škody simulujeme z exponenciálního rozdělení s parametrem $\tilde{\mu} = \frac{1}{2*365}$. Parametry obou rozdělení jsme převzali z Korn (2016). Ze stejného zdroje jsme se inspirovali při výběru kalendářních dat, konkrétně datum vyhodnocení T stanovíme na 31. 12. 2023, počátek platnosti smlouvy může nastat mezi 1. 1. 2015 a 31. 12. 2022. Délka smlouvy může být 1, 2 nebo 3 roky.

Datum 1. 1. 2015 si stanovíme jako den 0 a každému následujícímu datu přiřadíme přirozené číslo, tak jak jsou seřazeny v kalendáři za sebou. Tedy například jako den 10 chápeme 11. 1. 2015, jako den 365 chápeme 1. 1. 2016.

1) Údaje o smlouvách

Pro každou pojistnou smlouvu A_i, $i = 1, 2, \ldots, 1000$, získáme počátek její platnosti DZ_i jako realizace diskrétního rovnoměrného rozdělení nabývajícího hodnot 0, 1, 2, . . . , 2921. Hodnota 2921 odpovídá rozdílu dat 31. 12. 2022 a 1. 1. 2015.

Délku platnosti smlouvy A_i simulujeme z diskrétního rovnoměrného rozdělení nabývajícího hodnoty 365, $2 * 365$, $3 * 365$. Datum konce platnosti DK_i tedy získáme přičtením této realizace k začátku platnosti DZ_i. Takto dostaneme data konce a počátku platnosti pojistných smluv. Pro smlouvy $A_1, \ldots, A_5$ máme tyto hodnoty uvedeny v tabulce 4.1.

Index smlouvy	Počátek platnosti	Konec platnosti	Počet nahlá-šených škod
1	24. 2. 2019	24. 2. 2020	0
2	15. 4. 2018	14. 4. 2020	2
3	5. 7. 2015	4. 7. 2017	1
4	11. 12. 2015	10. 12. 2017	0
5	13. 5. 2022	12. 4. 2024	0

Tabulka 4.1: Údaje o smlouvách $A_1, \ldots, A_5$.

2) Údaje o škodách

Nyní chceme vygenerovat údaje o nahlášených škodách, konkrétně datum nastání a datum nahlášení. Pro smlouvu A_i určíme rizikovou expozici e_i podle délky působení pojistné smlouvy, tedy jako $(\min(DK_i,T) - DZ_i)/365$. Pro každou smlouvu A_i simulujeme celkový počet škod n_i z negativně binomického rozdělení s parametry k a $p/(p + e_i - pe_i)$.

Pro každou smlouvu A_i takovou, že $n_i > 0$, vygenerujeme údaje pro škodu $S_{i,j}$, $j \in$ 1, 2,..., n_i, konkrétně její datum nastání a datum nahlášení. Datum nastání $DN_{i,j}$ získáme přičtením horní celé části realizace z rovnoměrného rozdělení na intervalu $(0, \min(DK_i,T) - DZ_i)$ k DZ_i. Dobu zpoždění nahlášení $y_{i,j}$ simulujeme z exponenciálního rozdělení s parametrem $\mu = \frac{1}{2*365}$. Pak datum nahlášení $DM_{i,j}$ získáme přičtením horní celé části $y_{i,j}$ k $DN_{i,j}$.

Ze všech škod vybereme ty škody $S_{i,j}$, pro které $DM_{i,j} < T$. V našem případě jsme dostali soubor nahlášených škod o délce 720. Počet nahlášených škod na smlouvě A_i získáme jako $x_i = \sum_{j=1}^{n_i} \mathbb{I}_{DM_{i,j}<T}$ v případě kdy $n_i > 0$, pro případy kdy $n_i = 0$ je i $x_i = 0$. Hodnoty x_i pro smlouvy $A_1, \ldots, A_5$ jsou uvedeny v posledním sloupci tabulky 4.1. Údaje prvních pěti nahlášených škod $S_{i,j}$ (s nejnižšími možnými indexy i a j) jsou uvedeny v tabulce 4.2.

Index smlouvy	Index škody	Datum nastání	Datum nahlášení
2	1	9. 9. 2019	6. 4. 2022
2	2	17. 5. 2019	4. 3. 2023
3	1	10. 4. 2016	2. 12. 2016
6	1	31. 7. 2020	28. 3. 2022
7	1	2. 2. 2020	21. 6. 2020

Tabulka 4.2: Údaje o nahlášených škodách.

4.2 Odhad rozdělení počtu nenahlášených škod

Po simulaci údajů o smlouvách a škodách můžeme přejít na část, ve které budeme odhadovat parametry rozdělení doby zpoždění a rozdělení počtu škod nahlášených do data vyhodnocení T.

3) Rozdělení doby zpoždění

Předpokládáme, že rozdělení doby zpoždění se řídí exponenciálním rozdělením. Nahlášeným škodám, tedy škodám $S_{i,j}$, pro které $DM_{i,j} < T$, přiřadíme index u, $u = 1, 2, \ldots, 720$ v libovolném pořadí, stejné indexy přiřadíme i příslušným datům nastání a dobám zpoždění nahlášení. Maximálně věrohodný odhad $\widehat{\mu}$ dostaneme jako argument maxima logaritmické věrohodnostní funkce (2.4), kde $\boldsymbol{\theta} = \mu$, $r(x)$ je hustota exponenciálního rozdělení (2.3) a $R(x)$ je jeho distribuční funkce, tedy

$$\widehat{\mu} = \underset{\mu \in (0,\infty)}{\arg\max}\, \ell(\mu) = \underset{\mu \in (0,\infty)}{\arg\max} \sum_{u=1}^{720} \log\left(\frac{\mu e^{-\mu y_u}}{1 - e^{-\mu g_u}}\right),$$

kde hodnoty g_u získáme jako $T - DN_u$. Dostáváme $\widehat{\mu} \doteq 0.00126709$, což je odchylka -0.00010277 oproti hodnotě použité při simulaci $\tilde{\mu} = \frac{1}{2*365} \doteq 0.00136986$. Tedy střední hodnota odhadovaná ze simulovaných údajů je při zaokrouhlení na dny 789 dnů oproti skutečné střední hodnotě 730 dnů.

4) Pravděpodobnost nahlášení škody

K určení pravděpodobnosti q_i, $i = 1, 2, \ldots, 1000$ využijeme (2.1), tedy

$$q_i = \int_0^{v_i} \int_0^{z_i - w} r(y) \frac{1}{v_i} dy dw,$$

kde $r(y)$ je hustota exponenciálního rozdělení s parametrem $\widehat{\mu}$. Hodnoty z_i jsme získali jako $T - DZ_i$ a hodnoty v_i jako $(\min(DK_i,T) - DZ_i)$.

5) Rozdělení počtu nahlášených škod

Předpokládejme, že celkový počet škod na smlouvě s jednotkovou expozicí má negativně binomické rozdělení. Maximálně věrohodné odhady $\widehat{k}$ a $\widehat{p}$ dostáváme jako

$$(\widehat{k},\, \widehat{p})^\top = \operatorname*{arg\,max}_{\substack{k \in (0,\infty) \\ p \in (0,1)}} \ell(k,p),$$

kde $\ell(k,p)$ je určené rovností (1.11), za m jsme dosadili 1000. Dostáváme $\widehat{k} = 0.490336$ a $\widehat{p} = 0.490816$. Odchylka $\widehat{k}$ od $\tilde{k}$ je 0.00966372 a odchylka $\widehat{p}$ od $\tilde{p}$ je 0.00918384. Tedy střední hodnota rozdělení celkového počtu škod s odhadnutými parametry je 0.508686.

Nyní předpokládejme, že celkový počet škod se řídí Poissonovým rozdělením. Maximálně věrohodný odhad $\widehat{\lambda}$ získáme jako

$$\widehat{\lambda} = \operatorname*{arg\,max}_{\lambda \in (0,\infty)} \ell(\lambda),$$

kde $\ell(\lambda)$ je dáno rovností (1.9) a $m = 1000$. Ze simulovaných údajů dostáváme $\widehat{\lambda} = 0.501218$.

6) Rozdělení počtu nenahlášených škod

Pro smlouvu A_i, $i = 1, 2, \ldots, 1000$, dostaneme odhad parametrů rozdělení pro počet nastalých dosud nenahlášených škod dosazením $\widehat{k}$ a $\widehat{p}$ do (3.4) za k a p, respektive $\widehat{\lambda}$ do (3.3) za λ.

Jako možnost pro porovnání kvality odhadů uvádí Landry a Martin (2022) následující postup. Pro každé $j = 0, 1, \ldots$ porovnáme průměrnou pravděpodobnost, že na smlouvě bylo nahlášeno j škod, $\overline{Q(j)}$, a relativní četnost smluv, na kterých bylo nahlášeno j škod, $T(j)$, tedy

$$\overline{Q(j)} = \frac{1}{1000} \sum_{i=1}^{1000} P(X_i = j) \text{ a } T(j) = \frac{1}{1000} \sum_{i=1}^{1000} \mathbb{I}_{x_i = j}.$$

Rozdělení X_i je negativně binomické rozdělení s parametry $\widehat{k}$ a $\widehat{p}/(\widehat{p} + e_i q_i - \widehat{p} e_i q_i)$ nebo Poissonovo rozdělení s parametrem $\widehat{\lambda} e_i q_i$. Následuje tabulka 4.3, ve které jsou v každém sloupci pro hodnotu $j = 0, 1, \ldots, 6$ postupně hodnoty $T(j)$, $\overline{Q(j)}$ pro negativně binomický případ a $\overline{Q(j)}$ pro Poissonův případ.

	0	1	2	3	4	5	6
pozorované	0.597	0.165	0.104	0.0600	0.0240	0.0190	0.0120
neg. binom.	0.665	0.172	0.0736	0.0370	0.0203	0.0118	0.00716
Poisson	0.522	0.305	0.121	0.0387	0.0105	0.00244	0.000496

Tabulka 4.3: Hodnoty $T(j)$ a $\overline{Q(j)}$, pro $j = 0, 1, \ldots, 6$.

Hodnoty $\overline{Q(j)}$ a $T(j)$ jsou pro $j > 0$ velmi blízké nule. Proto je lepší porovnávat průměrné podmíněné pravděpodobnosti, že na smlouvě bylo nahlášeno j škod a podmíněné relativní četnosti smluv, na kterých bylo nahlášeno j škod za podmínky, že $j > 0$. $\overline{Q(j)}$ splňuje všechny vlastnosti pravděpodobnostní funkce, tedy můžeme uvažovat náhodnou veličinu Y určenou touto funkcí, ve smyslu $P(Y = j) = \overline{Q(j)}$, $j = 0, 1, 2 \dots$. Z definice podmíněné pravděpodobnosti pro $j > 0$

$$P(Y = j \mid Y > 0) = \frac{P(Y = j \cap Y > 0)}{P(Y > 0)} = \frac{P(Y = j)}{1 - P(Y = 0)} = \frac{\overline{Q(j)}}{1 - \overline{Q(0)}}.$$

Následuje tabulka 4.4, ve které jsou v každém sloupci pro hodnotu $j = 0, 1, \dots, 6$ postupně hodnoty $T(j)/(1 - T(0))$, $\overline{Q(j)}/(1 - \overline{Q(0)})$ pro negativně binomický případ a $\overline{Q(j)}/(1 - \overline{Q(0)})$ pro Poissonův případ.

	1	2	3	4	5	6	7
pozorované	0.41	0.258	0.149	0.060	0.047	0.0298	0.0174
neg. binom.	0.51	0.220	0.110	0.061	0.035	0.0214	0.0133
Poisson	0.64	0.253	0.081	0.0219	0.0051	0.00104	0.000187

Tabulka 4.4: Hodnoty $T(j)/(1 - T(0))$, $\overline{Q(j)}/(1 - \overline{Q(0)})$ pro $j = 1, 2, \dots, 7$.

Spočteme průměr počtu pozorovaných škod s přihlédnutím k rizikové expozici. Uvažujeme tedy, že na smlouvě A_i nastalo n_i/e_i škod. Průměr počtu škod P můžeme určit jako $P = \frac{1}{1000}\sum_{i=1}^{1000} n_i/e_i$, ze simulovaných údajů $P = 0.510254$. Tuto hodnotu porovnáme se středními hodnotami $\widehat{\lambda} = 0.501218$ a $\widehat{k}(1 - \widehat{p})/\widehat{p} = 0.508686$. Vidíme, že model s negativně binomickým rozdělením má menší odchylku než model s Poissonovým rozdělením, což bylo možné očekávat, vzhledem k tomu, že celkový počet škod byl generován z negativně binomického rozdělení. V tabulce 4.3 se hodnoty $\overline{Q(j)}$ pro negativně binomický model značně neliší od $T(j)/(1 - T(0))$ naproti tomu v tabulce 4.4 se hodnoty $\overline{Q(j)}/(1 - \overline{Q(0)})$ pro Poissonův model významně liší od $T(j)$ pro $j = 4, 5, \dots, 7$.

Závěr

V teoretické části práce jsme popsali modely pro odhad počtu nastalých dosud nenahlášených škod. V první kapitole jsme odvodili rozdělení pro počet nahlášených škod a zaobírali jsme se tím, jak celkový počet škod závisí na rizikové expozici, dále jsme si ukázali některé způsoby jejího stanovení. V druhé kapitole jsme popsali, jak určit pravděpodobnost, že škoda, která na smlouvě nastala, byla do data vyhodnocení nahlášena. Třetí kapitola obsahovala odvození rozdělení doby zpoždění.

V poslední kapitole jsme se věnovali simulační studii. Nejprve jsem simulovali údaje o smlouvách a nahlášených škodách, které bychom měli v praxi dostupné. Uvažovali jsme, že soubor smluv má délku 1000, celkový počet škod při jednotkové expozici jsme generovali z negativně binomického rozdělení s parametry 0,5 a 0,5 a rozdělení doby zpoždění se řídilo exponenciálním rozdělením s parametrem $\frac{1}{2*365}$. Pro pojistné smlouvy jsme vygenerovali data jejich konce a počátku platnosti, na základě délky působení pojistné smlouvy jsme stanovili její rizikovou expozici, a tak jsme mohli vygenerovat celkový počet škod na smlouvě. V příslušném počtu jsme generovali údaje pro škody, konkrétně datum nastání a datum nahlášení. Z těchto škod jsme vybrali ty, jejichž datum nahlášení předcházelo datu vyhodnocení. Z takto připravených údajů jsme odhadovali rozdělení doby zpoždění a rozdělení počtu nenahlášených škod. Za předpokladu, že rozdělení doby zpoždění má exponenciální rozdělení, jsme použili metodu maximální věrohodnosti pro odhad parametru tohoto rozdělení. Tento odhad jsme využili k určení pravděpodobností q_i. Opět jsme použili metodu maximální věrohodnosti pro odhad parametrů rozdělení počtu nahlášených škod, jak za předpokladu, že má negativně binomické rozdělení, tak za předpokladu, že se řídí Poissonovým rozdělením. Oba odhady jsme mezi sebou porovnali.

Seznam použité literatury

DeGroot, M. H. a Schervish, M. J. (2012). *Probability and statistics*. Pearson, London, 4th edition. ISBN 978-0-321-50046-5.

Korn, U. (2016). A Comprehensive, Non-aggregated, Stochastic Approach to Loss Development. *Variance*, **10**(1), 13 – 33.

Landry, D. D. a Martin, S. (2022). Policy-level unreported frequency model for pure ibnr estimation. *Variance*, **15**(1).

Zvára, K. (1997). *Pravděpodobnost a matematická statistika*. Matfyzpress, Praha, 1st edition. ISBN 80-85863-24-3.