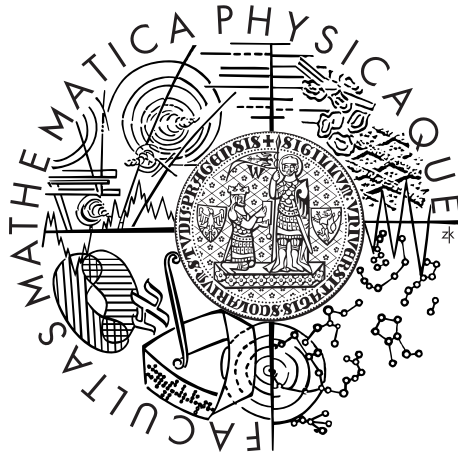


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Veronika Veselá

Shlukové bodové procesy v pojistné matematice

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Zbyněk Pawlas, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2012

Děkuji svému vedoucímu RNDr. Zbyňku Pawlasovi, Ph.D., za odborné vedení při psaní této práce, jeho rady, informace a vše, co jsem se při psaní této práce naučila.

Děkuji také RNDr. Ing. Radku Moravcovi a Mgr. Vendule Uchytlové za pomoc při hledání reálných dat.

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

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 v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne

Veronika Veselá

Název práce: Shlukové bodové procesy v pojistné matematice
Autor: Veronika Veselá
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: RNDr. Zbyněk Pawlas, Ph.D.

Abstrakt: V této práci se věnujeme bodovým procesům a jejich významu v pojistné matematice. Pomocí shlukových a kótovaných bodových procesů lze popsat model, který uvažuje doby vzniku pojistné události a doby a výšky příslušných úhrad. Zkoumáme dva konkrétní modely, které lze použít na výpočet predikce budoucích plateb a počtů plateb za pojistné události, které již nastaly. Prvním modelem je chain ladder v podobě Mackova modelu, u kterého ukazujeme odhady vývojových faktorů, rozptylu a jejich vlastnosti. Určujeme predikci o jeden a více kroků, na jejím základě pak vypočítáváme predikci rezervy. Krátce se také věnujeme asymptotickým vlastnostem. Druhý model je Poissonův shlukový model, kde nejdříve definujeme tento model a veličiny, které do něho vstupují. Posléze se věnujeme predikci o jeden a více kroků. Zajímá nás taky predikce při specifických rozděleních pro některé náhodné veličiny modelu. Na závěr aplikujeme obě metody predikce na simulovaná data a porovnáme jejich průměrné relativní absolutní chyby.

Klíčová slova: bodový proces, shlukový bodový proces, Mackův model, Poissonův shlukový model.

Title: Cluster point processes in insurance mathematics
Author: Veronika Veselá
Department: Department of Probability and Mathematical Statistics
Supervisor: RNDr. Zbyněk Pawlas, Ph.D.

Abstract: In the present work we study point processes and their importance in insurance mathematics. With the help of cluster and marked point processes we can describe a model that considers times of claim occurrence and times and heights of corresponding payments. We study two specific models which can be used to predict how much money is needed for claims which happened. The first model is chain ladder in the form of Mack's model. For this model we show chain ladder estimators of development factors, estimates of their variance and their properties. We try to find one-step ahead prediction and multi-step ahead prediction, which we use for calculating prediction of reserves. We shortly review asymptotic properties of the estimators in Mack's model. The second model is the Poisson cluster model. Firstly we define this model and the variables entering the model. Then we devote attention to one-step ahead and multi-step ahead prediction. We also study prediction when some variables have specific distributions. Finally, we use both methods of prediction on simulated data and compare their average relative absolute errors.

Keywords: point process, cluster point process, Mack's model, Poisson cluster model.

Obsah

Úvod	2
1 Bodové procesy	3
1.1 Základní definice	3
1.2 Poissonův bodový proces	5
1.3 Shlukové bodové procesy	7
1.4 Kótované bodové procesy	7
1.5 Bodové procesy v pojistné matematice	8
1.5.1 Panjerova rekurzivní formule	8
1.5.2 Shlukový model	11
2 Chain ladder	13
2.1 Formulace modelu	13
2.2 Mackův model	14
2.2.1 Asymptotické vlastnosti metody chain ladder	15
2.2.2 Momenty chain ladder odhadů	17
2.2.3 Predikce	23
2.2.4 Výpočet predikce rezervy	26
2.2.5 Příklad	28
3 Poissonův shlukový model	30
3.1 Popis modelu	30
3.2 Predikce	31
3.2.1 Jednokroková predikce	31
3.2.2 Vícekroková predikce	34
3.2.3 Predikce pro (a,b)-rozdělení	35
3.2.4 Asymptotické chování predikce	37
4 Simulace	40
4.1 Vysvětlení postupu	40
4.2 První experiment	41
4.3 Druhý experiment	43
4.4 Shrnutí	44
Závěr	46
Seznam použité literatury	48
Přílohy	49

Úvod

Každá neživotní pojišťovna potřebuje držet určité množství finančních prostředků na škody, které již nastaly, ale dosud nebyly nahlášeny nebo již byly nahlášeny, ale dosud nebyly zlikvidovány. Proto je pro každou takovou společnost velmi důležité, aby byla schopna realisticky odhadnout výši takovýchto prostředků. V této práci se budeme zabývat metodami, pomocí kterých můžeme odhadnout, jaký objem finančních prostředků bude ještě zapotřebí na likvidaci již nastalých pojistných událostí.

Nejnámější a v praxi pravděpodobně nejpoužívanější metodou dnešní doby v neživotních pojišťovnách je chain ladder, konkrétně metoda založená na stochastickém Mackově modelu, který byl poprvé prezentován v [3]. Avšak existují i jiné metody, jejichž pomocí lze získat podobné výsledky. Nás konkrétně bude zajímat nedávno vydaný článek [2] pojednávající o Poissonově shlukovém modelu. V této práci budeme studovat oba modely a pokusíme se porovnat jejich výhody a nevýhody.

S každou vzniklou pojistnou událostí je spojen shluk plateb, které pojišťovna v budoucnu vyplatí pojištěnému. Proto je přirozené tuto situaci modelovat pomocí shlukových bodových procesů. Jelikož nás zajímají i výše jednotlivých plateb, využijeme kótované bodové procesy. Tento obecný přístup pokrývá jak Mackův model, tak Poissonův shlukový model.

V první kapitole se zaměříme na bodové procesy. Uvedeme základní definice, ve kterých vysvětlíme, co to je bodový proces, Poissonův bodový proces, shlukové bodové procesy, kótované bodové procesy a také si nastíníme použití bodových procesů v pojistné matematice.

Ve druhé kapitole se podíváme na model chain ladderu, Mackův model a jeho základní vlastnosti jako například asymptotické chování či momenty odhadů. Odvodíme odhady vývojových faktorů, rozptylu a jejich vlastnosti. Řekneme si, jak se dá vytvářet predikce výše pojistných plnění na základě již známých dat o jeden a více kroků. Na základě této predikce zkonstruujeme predikci rezervy na pojistná plnění a určíme její relativní absolutní chybu. Na závěr této kapitoly budeme demonstrovat zmíněné postupy na reálných datech.

Třetí kapitola je věnována Poissonovu shlukovému modelu, na jehož základě se dají také provádět predikce budoucích úhrnů počtů plateb a úhrnů plateb za již nastalé pojistné události. Nejdříve vysvětlíme strukturu tohoto modelu. Ukážeme, jak vypadají jednokrokové a víceokrokové predikce. Posléze se zaměříme na predikce pro určitá rozdělení. Na závěr této kapitoly řekneme něco o asymptotickém chování predikce.

V poslední, čtvrté, kapitole provedeme simulaci dat a aplikaci těchto dvou zmíněných metod výpočtu predikce. Porovnáme je podle jejich průměrných relativních absolutních chyb a shrneme výsledky, ke kterým dojdeme.

1. Bodové procesy

V této kapitole se budeme věnovat základům z teorie bodových procesů. Budeme vycházet hlavně z knihy [6], v podkapitolách 1.3 a 1.4 využijeme také skripta [7].

1.1 Základní definice

Definice 1.1.1. (*Pravděpodobnostní prostor*)

Trojici $(\Omega, \mathcal{F}, \mathbb{P})$ nazveme pravděpodobnostním prostorem, jestliže Ω je neprázdná množina, \mathcal{F} je σ -algebra podmnožin množiny Ω a \mathbb{P} je pravděpodobnostní míra na \mathcal{F} .

Nechť E je borelovská podmnožina konečně dimenzionálního euklidovského prostoru \mathbb{R}^d a je vybavena σ -algebrou $\mathcal{E} = \mathcal{B}(E)$ borelovských množin z E . Lebesgueovu míru množiny $A \in \mathcal{E}$ budeme značit $|A| = \text{Leb}(A)$. Dále zavedeme Diracovu míru ε_x v $x \in E$:

$$\varepsilon_x(A) = I_A(x) = \begin{cases} 1 & \text{pro } x \in A, \\ 0 & \text{pro } x \notin A, \end{cases} \quad A \in \mathcal{E}.$$

Uvažujme posloupnost (X_i) náhodných vektorů v E a pro $A \in \mathcal{E}$ definujme

$$N(A) = \# \{i \geq 1 : X_i \in A\},$$

tj. $N(A)$ sčítá počet X_i spadajících do A . Znamená to, že $N(A) = N(A, \omega)$ je náhodné pro danou množinu A a pro pevné $\omega \in \Omega$ definuje $N(\cdot, \omega)$ čítací míru na σ -algebře \mathcal{E} takovou, že její atomy jsou X_i . Míra N , která může být zapsána ve tvaru $N = \sum_{i \geq 1} \varepsilon_{X_i}$, je bodový proces, pokud posloupnost (X_i) d -dimenzionálních náhodných vektorů je taková, že s pravděpodobností 1 libovolná kompaktní množina $K \in \mathcal{E}$ obsahuje pouze konečný počet bodů X_i . Pak, pro skoro všechna $\omega \in \Omega$,

$$N(A, \omega) = \sum_{i \geq 1} \varepsilon_{X_i(\omega)}(A), \quad A \in \mathcal{E},$$

definuje bodovou míru na \mathcal{E} .

Obecně pro danou posloupnost $(x_i)_{i \geq 1}$ v E ,

$$m(A) = \sum_{i \geq 1} \varepsilon_{x_i}(A) = \# \{i \geq 1 : x_i \in A\}, \quad A \in \mathcal{E},$$

definuje čítací míru na \mathcal{E} , která se nazývá *bodová míra*, jestliže $m(K) < \infty$ pro všechny kompaktní množiny $K \in \mathcal{E}$. Toto znamená, že žádná kompaktní množina K nesmí obsahovat nekonečně mnoho bodů x_i . Nechť $M_p(E)$ je prostor všech bodových měr na E vybavený nejmenší σ -algebrou $\mathcal{M}_p(E)$, která obsahuje všechny množiny tvaru

$$U_{A,k} = \{m \in M_p(E) : m(A) = k\} \tag{1.1}$$

pro $A \in \mathcal{E}$ a $k \in \mathbb{N} \cup \{0, \infty\}$, tj. je to nejmenší σ -algebra zaručující, že zobrazení $m \mapsto m(A)$ je měřitelné pro libovolné $A \in \mathcal{E}$.

Předpokládejme, že $m = \sum_{i \geq 1} \varepsilon_{x_i}$ je bodová míra na E . Necht (y_i) je podposloupnost (x_i) obsahující vzájemně různé hodnoty x_i (bez opakování). Definujme *multiplicitu* y_i jako $n_i = \#\{j \geq 1 : y_i = x_j\}$. Pak můžeme psát $m = \sum_{i \geq 1} n_i \varepsilon_{y_i}$. Jestliže $n_i = 1$ pro všechna i , pak m se nazývá *jednoduchá bodová míra*. Prostor všech jednoduchých bodových měr na E označme $M_s(E)$.

Definice 1.1.2. (*Bodový proces*)

Bodový proces N se stavovým prostorem E je měřitelné zobrazení z $(\Omega, \mathcal{F}, \mathbb{P})$ do $(M_p(E), \mathcal{M}_p(E))$. Řekneme, že bodový proces N je *jednoduchý*, jestliže nabývá hodnot z $M_s(E)$ s pravděpodobností 1. Bodový proces je *konečný*, pokud $N(E) < \infty$ s.j.

Jinými slovy, bodový proces N je náhodný element, který nabývá bodové míry jako hodnoty: pro každé $\omega \in \Omega$ je hodnota $m(\cdot) = N(\cdot, \omega)$ bodová míra. Speciálně platí, že $N(K) < \infty$ pro kompaktní množiny $K \in \mathcal{E}$. V kontextu čítací míry je přirozené, že $N(A)$ může také nabývat hodnotu ∞ pro nekompaktní množiny A .

Jestliže realizace bodového procesu N jsou jednoduché bodové míry s pravděpodobností 1, pak N je *jednoduchý bodový proces*. Bodový proces N s reprezentací $N = \sum_{i \geq 1} \varepsilon_{X_i}$ pro posloupnost náhodných vektorů X_i s hodnotami v E je jednoduchý, jestliže všechny body X_i , $i = 1, 2, \dots$ jsou navzájem různé s pravděpodobností 1.

Další výsledek dává ospravedlnění skutečnosti, že můžeme interpretovat bodový proces N jako soubor $(N(A))_{A \in \mathcal{E}}$ náhodných veličin $N(A)$ s hodnotami v $\{0, 1, \dots, \infty\}$.

Lemma 1.1.3. (*Bodový proces jako soubor náhodných čítacích veličin*)

Zobrazení N z $(\Omega, \mathcal{F}, \mathbb{P})$ do $(M_p(E), \mathcal{M}_p(E))$ je bodový proces na E právě tehdy, když pro každé $A \in \mathcal{E}$ je $N(A)$ náhodná veličina s hodnotami v $\mathbb{N} \cup \{0, \infty\}$ taková, že $N(A) < \infty$ s.j., pokud A je kompaktní.

Důkaz. Necht N je bodový proces na E . Z definice bodového procesu plyne, že zobrazení $N : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (M_p(E), \mathcal{M}_p(E))$ je měřitelné. Pro danou borelovskou množinu $A \in \mathcal{E}$ uvažujme zobrazení $f_A : (M_p(E), \mathcal{M}_p(E)) \rightarrow (\mathbb{N} \cup \{0, \infty\}, \mathcal{B}(\mathbb{N} \cup \{0, \infty\}))$ dané předpisem $f_A(m) = m(A)$. Toto zobrazení je měřitelné, což plyne ze skutečnosti, že σ -algebra $\mathcal{M}_p(E)$ je generována množinami (1.1). Složené zobrazení $N(A, \omega) = f_A(N(\cdot, \omega))$ je pak také měřitelné, a proto $N(A)$ je náhodná veličina, a z definice bodového procesu víme, že $N(A)$ je konečná pro kompaktní A .

Nyní dokážeme opačnou implikaci. Necht $N(A)$ je soubor náhodných veličin s hodnotami v $\mathbb{N} \cup \{0, \infty\}$ a $N(A) < \infty$, pokud A je kompaktní. Pak vzor množiny $U_{A,k}$ při zobrazení N je $N^{-1}(U_{A,k}) = \{\omega \in \Omega : N(A, \omega) = k\}$, což je vzor množiny $\{k\}$ při zobrazení $N(A)$, a proto je prvkem \mathcal{F} . Zjistili jsme, že $N^{-1}(U_{A,k}) \in \mathcal{F}$ pro každé $A \in \mathcal{E}$ a $k \in \mathbb{N} \cup \{0, \infty\}$. Protože systém množin $(U_{A,k})$ generuje σ -algebru $\mathcal{M}_p(E)$, je N bodový proces. \square

Definice 1.1.4. (*Míra intenzity*)

Necht N je bodový proces na E . Jeho míru intenzity definujeme jako $\mu(A) = \mathbb{E}N(A)$ pro $A \in \mathcal{E}$.

Příklad 1. (Proces obnovy definuje jednoduchý bodový proces)

Uvažujme proces obnovy (T_i) , tj. T_i jsou body z náhodné procházky s nezávislými, stejně rozdělenými, kladnými kroky velikosti Y_i :

$$T_0 = 0, \quad T_i = Y_1 + \cdots + Y_i, \quad i \in \mathbb{N}.$$

Ze silného zákona velkých čísel vyplývá, že $T_i \uparrow \infty$ s.j. pro $i \rightarrow \infty$, a proto realizace $(T_i(\omega))$ posloupnosti (T_i) nemají konečný počet bodů s pravděpodobností 1. Proto náhodná veličina

$$N(A) = \sum_{i=1}^{\infty} \varepsilon_{T_i}(A) = \#\{i \in \mathbb{N} : T_i \in A\}$$

je konečná s.j. pro každou omezenou borelovskou podmnožinu A prostoru hodnot $E = (0, \infty)$. Protože $Y_i = T_i - T_{i-1} > 0$ s.j., všechny body T_i jsou vzájemně různé s pravděpodobností 1, a proto N je jednoduchý bodový proces.

Příklad 2. (Složený bodový proces)

Nechť $\sum_{i \geq 1} \varepsilon_{T_i}$ je jednoduchý bodový proces na $(0, \infty)$. Předpokládáme, že (Z_i) je posloupnost nezávislých, stejně rozdělených, kladných náhodných veličin, nezávislých na (T_i) . Náhodný proces definovaný vztahem

$$S(t) = \sum_{i \geq 1} Z_i I_{[0,t]}(T_i), \quad t \geq 0, \quad (1.2)$$

vyjadřuje úhrn náhodných přírůstků Z_i , které nastaly do času t v náhodných okamžicích T_i . Následující náhodné veličiny definují bodový proces s prostorem hodnot $E = (0, \infty)^2$:

$$N(A) = \sum_{i \geq 1} \varepsilon_{(T_i, Z_i)}(A) = \#\{i \geq 1 : (T_i, Z_i) \in A\}.$$

Protože (T_i) jsou vzájemně různé, jsou také dvojice (T_i, Z_i) vzájemně různé a N je jednoduchý bodový proces.

1.2 Poissonův bodový proces

Poissonův proces, nebo taky jinak nazývaná Poissonova náhodná míra, představuje jednu z nejdůležitějších tříd bodových procesů. Vzniká přirozenou cestou jako limita „binomického bodového procesu“, tj. bodového procesu s nezávisle rozmístěnými body.

Připomeňme, že míra μ na E je Radonova míra, jestliže je konečná na kompaktních množinách $A \in \mathcal{E}$.

Definice 1.2.1. (Poissonův proces)

Nechť μ je Radonova míra na E . Bodový proces N se nazývá Poissonův proces nebo Poissonova náhodná míra s mírou intenzity μ (zapisujeme $PRM(\mu)$ z anglického Poisson random measure), jestliže jsou splněny následující dvě podmínky:

1. Náhodná veličina $N(A)$ má Poissonovo rozdělení s parametrem $\mu(A)$ pro libovolné $A \in \mathcal{E}$.

2. Náhodné veličiny $N(A_1), \dots, N(A_m)$ jsou vzájemně nezávislé pro libovolné po dvou disjunktní množiny $A_1, \dots, A_m \in \mathcal{E}$ a $m \in \mathbb{N}$.

Pro množiny A , pro které $\mu(A) = s$, kde $s \in \{0, \infty\}$, budeme používat konvenci, že $N(A) = s$ s.j. Toto je v souladu se vztahem $Y_\lambda \xrightarrow{P} s$ pro $\lambda \rightarrow s$, kde Y_λ je náhodná veličina s Poissonovým rozdělením s parametrem λ a $s \in \{0, \infty\}$.

Míra intenzity Poissonova procesu N splňuje $\mathbb{E}N(A) = \mu(A)$ pro všechny $A \in \mathcal{E}$. Z definice PRM vyplývá, že míra intenzity μ určuje rozdělení N . Toto je podobné faktu, že Poissonovo rozdělení je dáno střední hodnotou.

Radonova vlastnost pro míru intenzity μ nám zaručuje, že pro každou kompaktní množinu $K \in \mathcal{E}$ platí $\mathbb{E}N(K) = \mu(K) < \infty$, a proto $N(K) < \infty$ s.j., což bylo požadováno v definici bodového procesu.

Poissonův proces N s konečnou mírou intenzity μ , tj. $\mu(E) < \infty$, má reprezentaci:

$$N(A) = \sum_{i=1}^{\tau} I_A(X_i), \quad A \in \mathcal{E},$$

kde (X_i) je posloupnost nezávislých stejně rozdělených náhodných vektorů s hodnotami v E , nezávislá na náhodné veličině τ , která má Poissonovo rozdělení s parametrem $\mu(E)$, a X_i mají společné rozdělení

$$P(X_1 \in A) = \frac{\mu(A)}{\mu(E)}, \quad A \in \mathcal{E}.$$

Příklad 3. (Homogenní Poissonův proces na přímce)

Uvažujme Poissonův proces N na stavovém prostoru $E = [0, \infty)$ s mírou intenzity λLeb pro nějaké $\lambda > 0$, kde Leb značí Lebesgueovu míru na E . Definujme náhodný proces $N(t) = N([0, t])$, $t \geq 0$. Tento proces má stacionární přírůstky, protože pro každé $0 < a < b < \infty$ a $h > 0$ platí, že $N(a+h, b+h]$ má Poissonovo rozdělení s parametrem $\lambda(b-a)$. Navíc pro $0 = t_0 < t_1 < \dots < t_m < \infty$ jsou množiny $(t_{i-1}, t_i]$, $i = 1, \dots, m$ po dvou disjunktní, tudíž přírůstky $N(t_{i-1}, t_i]$, $i = 1, \dots, m$ jsou vzájemně nezávislé. Jelikož $\mathbb{E}N(0) = \mathbb{E}N(\{0\}) = \lambda|0| = 0$, dostáváme $N(0) = 0$ s.j. Předpokládáme-li ještě, že s pravděpodobností jedna jsou trajektorie $(N(t, \omega))_{t \geq 0}$ procesu N zprava spojité pro $t \geq 0$ a mají limity zleva pro $t > 0$, pak dostáváme vlastnosti homogenního Poissonova procesu $(N(t))_{t \geq 0}$ s intenzitou $\lambda > 0$. Z tohoto důvodu můžeme nazývat Poissonův proces na $[0, \infty)$ s mírou intenzity λLeb homogenním Poissonovým procesem na stavovém prostoru $[0, \infty)$.

Motivováni předchozím příkladem můžeme definovat homogenní Poissonův proces nebo homogenní PRM na borelovském stavovém prostoru $E \subset \mathbb{R}^d$ jako Poissonův proces s mírou intenzity $\lambda \text{Leb}(\cdot \cap E)$, číslo $\lambda > 0$ se nazývá intenzita.

Obecněji, jestliže míra intenzity μ Poissonova procesu je absolutně spojitá vzhledem k Lebesgueově míře, tj. existuje nezáporná funkce $\lambda(\cdot)$ taková, že

$$\mu(A) = \int_A \lambda(x) dx, \quad A \in \mathcal{E},$$

pak $\lambda(\cdot)$ je funkce intenzity Poissonova procesu.

Příklad 4. (Zúžení Poissonova procesu na menší stavový prostor je opět Poissonův proces)

Uvažujme Poissonův proces N na stavovém prostoru E s mírou intenzity μ a necht $E' \in \mathcal{E}$ je měřitelná podmnožina E . Definujme bodový proces

$$N'(A) = N(A), \quad A \in \mathcal{E}' = \mathcal{B}(E').$$

Pak z definice Poissonova procesu na E plyne, že $N'(A)$ má Poissonovo rozdělení s parametrem $\mu(A)$ pro $A \in \mathcal{E}'$ a $N'(A_1), \dots, N'(A_m)$ jsou nezávislé pro po dvou disjunktí množiny A_1, \dots, A_m z \mathcal{E}' . To znamená, že restrikce N' z N na borelovskou množinu $E' \subset E$ je znovu Poissonův proces s mírou intenzity μ' , která je restrikcí μ na \mathcal{E}' .

1.3 Shlukové bodové procesy

Definice 1.3.1. (Shlukový bodový proces)

Necht jsou dány (na jednom pravděpodobnostním prostoru) bodový proces N_R (tzv. rodičovský proces) a konečné bodové procesy N_x (tzv. dceřiné procesy) na E , $x \in E$. Potom bodový proces daný vztahem

$$N(A) = \int N_x(A) N_R(dx), \quad A \in \mathcal{E},$$

je shlukový bodový proces na E . Jsou-li shluky N_x navzájem nezávislé a nezávislé na N_R , nazýváme N procesem s nezávislými shluky.

Poznámka 1.3.2. Shlukový bodový proces tedy znamená, že každý bod rodičovského procesu generuje konečný shluk bodů.

Definice 1.3.3. (Poissonův shlukový bodový proces)

Shlukový bodový proces s nezávislými shluky takový, že rodičovský proces N_R je Poissonův bodový proces na E , se nazývá Poissonův shlukový bodový proces.

Pro $z \in \mathbb{R}^d$ označme t_z operátor posunutí na $M_p(E)$:

$$(t_z m)(A) = m(A - z), \quad A \in \mathcal{E},$$

kde $A - z = \{y - z : y \in A\}$ je posunutí množiny A o vektor $-z$.

Definice 1.3.4. (Neymanův-Scottové bodový proces)

Necht N je Poissonův shlukový bodový proces. Pokud centrované shluky $t_{-x}N_x$ jsou stejně rozdělené, nazývá se N Neymanův-Scottové bodový proces. Jestliže navíc jsou $t_{-x}N_x$ konečné Poissonovy bodové procesy, pak mluvíme o Neymanově-Scottové Poissonově procesu.

1.4 Kótované bodové procesy

Definice 1.4.1. (Kótovaný bodový proces) Necht $N = \sum_{i \geq 1} \varepsilon_{X_i}$ je jednoduchý bodový proces na E a (Z_i) je posloupnost náhodných elementů s hodnotami v měřitelném prostoru (E_M, \mathcal{E}_M) . Kótovaný bodový proces s prostorem kót E_M definujeme jako

$$N_M = \sum_{i \geq 1} \varepsilon_{(X_i, Z_i)}.$$

Kótovaný bodový proces N_M jsme obdrželi tak, že jsme každému bodu procesu N přiřadili určitou dodatečnou hodnotu (tzv. kótu).

Složený bodový proces z příkladu 2 lze chápat jako kótovaný bodový proces, kdy kóty jsou přírůstky Z_i nastávající v časech T_i . Prostor kót je $M = (0, \infty)$.

Každý shlukový bodový proces je generovaný kótovaným bodovým procesem, kde prostor kót je tvořený konečnými bodovými mírami. Konkrétně

$$N_M = \sum_{i \geq 1} \varepsilon_{(X_i, t - X_i) N_{X_i}}, \quad (1.3)$$

kde $N_R = \sum_{i \geq 1} \varepsilon_{X_i}$ je rodičovský proces a N_{X_i} jsou dceřiné procesy.

Definice 1.4.2. (*Nezávisle kótovaný bodový proces*)

Řekneme, že $N_M = \sum_{i \geq 1} \varepsilon_{(X_i, Z_i)}$ je nezávisle kótovaný bodový proces, jestliže posloupnost kót (Z_i) je posloupnost nezávislých, stejně rozdělených náhodných elementů v E_M a je nezávislá na bodovém procesu $N = \sum_{i \geq 1} \varepsilon_{X_i}$.

Nezávisle kótovaný bodový proces (1.3) takový, že $N = \sum_{i \geq 1} \varepsilon_{X_i}$ je Poissonův proces, generuje Neymanův-Scottův bodový proces.

1.5 Bodové procesy v pojistné matematice

Jedním z mnoha oborů, kde mají bodové procesy uplatnění, je pojistná matematika. Například na časy nahlášení pojistných událostí lze nahlížet jako na realizaci bodového procesu na $E = (0, \infty)$. S každou pojistnou událostí je spojeno pojistné plnění. To nás přirozeně vede ke složenému bodovému procesu z příkladu 2. Proces $S(t)$ definovaný vztahem (1.2) se obvykle nazývá *proces celkových pojistných nároků* (v angličtině: total claim process). V podpodkapitole 1.5.1 uvedeme numerický rekurzivní postup pro výpočet rozdělení náhodné veličiny $S(t)$ v případě, že velikosti škod nabývají diskrétních hodnot. V této práci se zaměříme na případy, kdy každá pojistná událost spouští proud plateb od pojistitele k pojištěnému. To lze modelovat pomocí shlukových bodových procesů jak bude vysvětleno v podpodkapitole 1.5.2.

1.5.1 Panjerova rekurzivní formule

V této části odvodíme metodu pro určení rozdělení úhrnu škod, který je dán vztahem (1.2). Budeme uvažovat, že $t \geq 0$ je pevné, tj. studujeme náhodnou veličinu $S(t)$ a ne náhodný proces $(S(t))_{t \geq 0}$. Potlačíme závislost $S(t)$ na t a píšeme $S = S(t)$. Tedy

$$S = \sum_{i=1}^N Z_i,$$

kde náhodná veličina N označuje počet pojistných událostí T_i do času t a je nezávislá na (Z_i) . Dále píšeme

$$S_0 = 0, \quad S_n = Z_1 + \cdots + Z_n, \quad n \in \mathbb{N},$$

pro proces náhodných částečných součtů (náhodná procházka) generovaných škodami velikosti Z_i .

Distribuční funkce S je dána výrazem

$$P(S \leq x) = \mathbb{E}[P(S \leq x|N)] = \sum_{n=0}^{\infty} P(S_n \leq x)P(N = n), \quad x \geq 0.$$

Z tohoto tvaru vidíme, že úhrn škod S má komplikovanou strukturu, proto k přesnému určení rozdělení S použijeme numerickou metodu, která se většinou nazývá *Panjerova rekurzivní formule*.

Začneme se základními předpoklady:

(P1) Velikosti škod Z_i nabývají hodnot v \mathbb{N}_0 .

(P2) Počet škod N má rozdělení typu

$$q_n = P(N = n) = \left(a + \frac{b}{n}\right) q_{n-1}, \quad n \in \mathbb{N}$$

pro nějaké $a, b \in \mathbb{R}$.

Podmínka (P2) se nazývá (a, b) -podmínka, respektive rozdělení splňující tuto podmínku tvoří (a, b) -třídou.

Tři standardní rozdělení splňují (a, b) -podmínku:

1. Poissonovo rozdělení s intenzitou λ , $\text{Pois}(\lambda)$, kde $a = 0$, $b = \lambda \geq 0$. V tomto případě obdržíme tzv. (a, b) -oblast

$$R_{\text{Pois}} = \{(a, b) : a = 0, b \geq 0\}.$$

2. Binomické rozdělení $\text{Bin}(n, p)$, kde $a = -\frac{p}{1-p} < 0$, $b = -a(n+1)$, $n \in \mathbb{N}_0$. V tomto případě obdržíme (a, b) -oblast

$$R_{\text{Bin}} = \{(a, b) : a < 0, b = -a(n+1) \text{ pro nějaké } n \in \mathbb{N}_0\}.$$

3. Negativně binomické rozdělení s parametry (p, ν) :

$$q_n = \binom{\nu + n - 1}{n} n^\nu (1-p)^n, \quad n \in \mathbb{N}_0, \quad p \in (0, 1), \quad \nu > 0,$$

kde $0 < a = 1-p < 1$, $b = (1-p)(\nu-1)$ a $a+b > 0$. V tomto případě obdržíme (a, b) -oblast

$$R_{\text{Negbin}} = \{(a, b) : 0 < a < 1, a+b > 0\}.$$

Tyto tři rozdělení jsou jediné rozdělení na \mathbb{N}_0 splňující (a, b) -podmínku.

Nyní formulujeme Panjerovu rekurzivní formuli jako větu.

Věta 1.5.1. (*Panjerova rekurzivní formule*)

Předpokládejme podmínky (P1) a (P2) pro rozdělení Z_i a N . Pak pravděpodobnosti $p_n = P(S = n)$ můžou být spočítány rekurzivně následujícím způsobem:

$$p_0 = \begin{cases} q_0, & \text{jestliže } P(Z_1 = 0) = 0, \\ \mathbb{E}[P(Z_1 = 0)]^N & \text{jinak,} \end{cases}$$

$$p_n = \frac{1}{1 - aP(Z_1 = 0)} \sum_{i=1}^n \left(a + \frac{bi}{n}\right) P(Z_1 = i)p_{n-i}, \quad n \in \mathbb{N}.$$

Poznámka 1.5.2. Protože parametr a je nutně menší než 1, všechny vzorce p_n jsou dobře definovány.

Důkaz. Začneme s

$$p_0 = P(N = 0) + P(S = 0, N > 0).$$

Posledně jmenovaný vztah se rovná q_0 , jestliže $P(Z_1 = 0) = 0$. Jinak platí

$$\begin{aligned} p_0 &= q_0 + \sum_{i=1}^{\infty} P(Z_1 = 0, \dots, Z_i = 0)P(N = i) \\ &= q_0 + \sum_{i=1}^{\infty} [P(Z_1 = 0)]^i P(N = i) \\ &= \mathbb{E}[P(Z_1 = 0)]^N. \end{aligned}$$

Nyní se vraťme k případu p_n , $n \in \mathbb{N}$. Věta o úplné pravděpodobnosti a (a, b) -podmínka vedou k následujícímu výrazu

$$p_n = \sum_{i=1}^{\infty} P(S_i = n)q_i = \sum_{i=1}^{\infty} P(S_i = n) \left(a + \frac{b}{i}\right) q_{i-1}. \quad (1.4)$$

Poznamenejme, že

$$\begin{aligned} \mathbb{E}\left(a + \frac{bZ_1}{n} \middle| S_i = n\right) &= \mathbb{E}\left(a + \frac{bZ_1}{Z_1 + \dots + Z_i} \middle| S_i = n\right) \\ &= a + \frac{b}{i}, \end{aligned} \quad (1.5)$$

protože díky tomu, že X_i jsou nezávislé, stejně rozdělené, platí

$$1 = \mathbb{E}\left(\frac{S_i}{S_i} \middle| S_i\right) = \sum_{k=1}^i \mathbb{E}\left(\frac{Z_k}{S_i} \middle| S_i\right) = i\mathbb{E}\left(\frac{Z_1}{S_i} \middle| S_i\right).$$

Pozorujeme, že

$$\begin{aligned} \mathbb{E}\left(a + \frac{bZ_1}{n} \middle| S_i = n\right) &= \sum_{k=0}^n \left(a + \frac{bk}{n}\right) P(Z_1 = k | S_i = n) \\ &= \sum_{k=0}^n \left(a + \frac{bk}{n}\right) \frac{P(Z_1 = k, S_i - Z_1 = n - k)}{P(S_i = n)} \\ &= \sum_{k=0}^n \left(a + \frac{bk}{n}\right) \frac{P(Z_1 = k)P(S_{i-1} = n - k)}{P(S_i = n)}. \end{aligned} \quad (1.6)$$

Dosadíme-li (1.5) a (1.6) do (1.4) a přehodíme-li pořadí sčítání, dostaneme:

$$\begin{aligned} p_n &= \sum_{i=1}^{\infty} \sum_{k=0}^n \left(a + \frac{bk}{n}\right) P(Z_1 = k)P(S_{i-1} = n - k)q_{i-1} \\ &= \sum_{k=0}^n \left(a + \frac{bk}{n}\right) P(Z_1 = k) \left[\sum_{i=1}^{\infty} P(S_{i-1} = n - k)q_{i-1} \right] \\ &= \sum_{k=0}^n \left(a + \frac{bk}{n}\right) P(Z_1 = k)P(S = n - k) \\ &= \sum_{k=0}^n \left(a + \frac{bk}{n}\right) P(Z_1 = k)p_{n-k}. \end{aligned}$$

Na závěr obdržíme

$$p_n = aP(Z_1 = 0)p_n + \sum_{k=1}^n \left(a + \frac{bk}{n}\right) P(Z_1 = k)p_{n-k},$$

což dává konečný výsledek pro p_n . □

1.5.2 Shlukový model

Budeme uvažovat shlukový bodový proces na $E = (0, \infty)$ takový, že rodičovský proces $N_R = \sum_{i \geq 1} \varepsilon_{T_i}$ je jednoduchý a dceřiné procesy jsou tvořeny událostmi T_{ij} následujícími po příslušném rodičovském bodě T_i . Čas T_i interpretujeme jako čas vzniku pojistné události a body dceřiného procesu T_{ij} jako čas j -té platby i -té události. U každé platby nás ještě zajímá její výše, kterou budeme značit pomocí X_{ij} .

Definice 1.5.3. (*Obecný shlukový model*)

Nechť $N_R = \sum_{i \geq 1} \varepsilon_{T_i}$ je jednoduchý bodový proces na $(0, \infty)$. Pro $i \geq 1$ položme

$$T_{ij} = T_i + \sum_{k=1}^j Y_{ik}, \quad j = 1, \dots, K_i,$$

kde $(Y_{ik})_{k \geq 1}$ je posloupnost kladných náhodných veličin a K_i je kladná celočíselná náhodná veličina. Nechť (X_{ik}) je systém kladných náhodných veličin, $i \geq 1, k \geq 1$. Časovým okamžikům T_i přiřadíme náhodné elementy

$$Z_i = ((Y_{ik})_{k \geq 1}, (X_{ik})_{k \geq 1}, K_i), \quad i \geq 1,$$

které nabývají hodnoty v množině

$$E_M = (0, \infty)^\infty \times (0, \infty)^\infty \times \mathbb{N},$$

kde $(0, \infty)^\infty$ označuje prostor posloupností s kladnými prvky:

$$(0, \infty)^\infty = \{(x_k)_{k \geq 1} : x_k \in (0, \infty)\}.$$

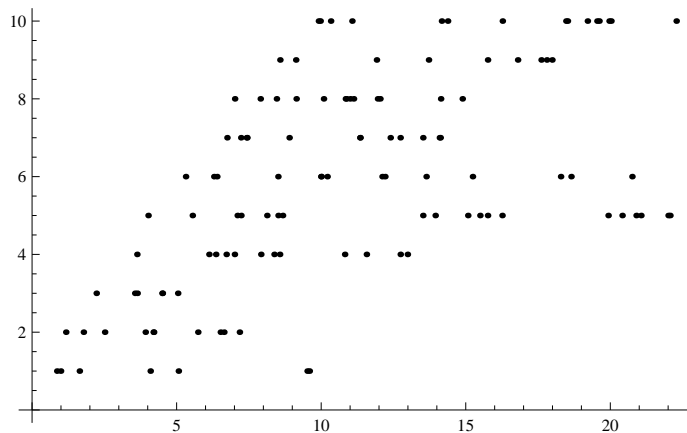
Dostáváme kótovaný bodový proces

$$N_M = \sum_{i \geq 1} \varepsilon_{(T_i, Z_i)} \tag{1.7}$$

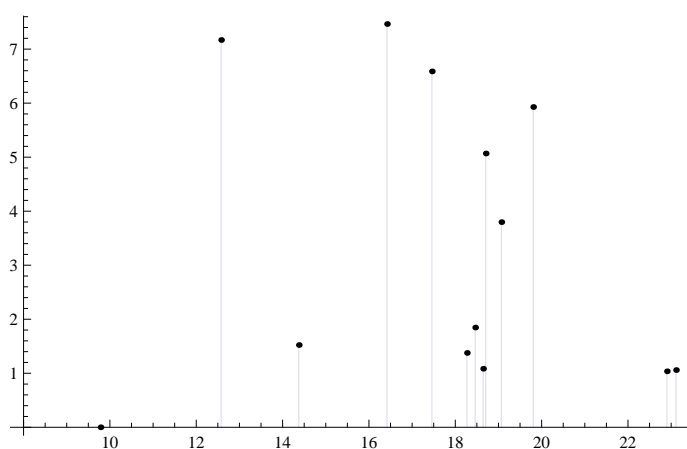
se stavovým prostorem $(0, \infty) \times E_M$.

Díky této definici je proces plateb i -té události popsán dvojicemi (T_{ij}, X_{ij}) , $j = 1, \dots, K_i$. Událost i se vypořádává v čase $T_{iK_i} = T_i + \sum_{k=1}^{K_i} Y_{ik}$ odpovídající částkou $\sum_{k=1}^{K_i} X_{ik}$.

Na obrázku 1.1 jednotlivé řádky znázorňují pojistnou událost a k ní příslušnou posloupnost plateb. Doba vzniku pojistné události je znázorněna první tečkou a následující tečky jsou již okamžiky jednotlivých plateb od pojistitele k pojištěnému. Obrázek 1.1 jsme vytvořili v programu Mathematica 8.0 [8] na základě předpokladu, že počty plateb K_i k dané události mají Poissonovo rozdělení, kde



Obrázek 1.1: Znázornění dob pojistných událostí a shluků příslušných plateb



Obrázek 1.2: Znázornění velikostí X_{ij} plateb v časech T_{ij}

jsme zvolili parametr 10, a časy Y_{ik} mezi jednotlivými platbami jsou nezávislé a mají exponenciální rozdělení, kde jsme položili parametr roven 1.

Na obrázku 1.2 první tečka zleva znamená dobu vzniku pojistné události a následující tečky označují čas a velikost jednotlivých plateb od pojistitele k pojištěnému. Tento obrázek jsme také vytvořili v programu Mathematica na základě předpokladu, že výše plateb jsou nezávislé a mají logaritmicko-normální rozdělení, parametry jsme zvolili 1 a 1, a doby mezi nimi mají exponenciální rozdělení s parametrem také 1.

2. Chain ladder

Model shlukového bodového procesu obsahuje chain ladder, který je používán většinou aktuáři k předpovědi počtu událostí a celkové požadované částky v budoucích letech na základě dat z minulých let.

2.1 Formulace modelu

Chain ladder je model, který je často používán v pojišťovnictví ke stanovení rezerv. Pro formulaci modelu je vhodné označit časové období (řekněme roky) jako

$$C_i = (i - 1, i], \quad i \in \mathbb{N},$$

a úhrn několika období jako

$$C_{ij} = C_i \cup \dots \cup C_{i+j} = (i - 1, i + j], \quad i \in \mathbb{N}, \quad j \in \mathbb{N}_0.$$

Data potřebná k výpočtu rezervy na pojistná plnění na základě minulého škodního vývoje se nejčastěji reprezentují ve formě *vývojových trojúhelníků*. Jedná se o schéma, v němž hodnoty v jednom řádku jsou vztaženy ke stejnému referenčnímu období (nejčastěji se jedná o rok, ve kterém došlo ke vzniku daných škod), data uvedená v jednom sloupci mají společný údaj o zpoždění v úhradě nebo v nahlášení. Vývojové trojúhelníky rozdělujeme na dva typy: *kumulativní a nekumulativní trojúhelníky*. V kumulativním trojúhelníku je vstupem hodnota vztahující se ke škodám uhrazeným nebo nahlášeným do určité doby od vzniku. V nekumulativním trojúhelníku se jedná o škody uhrazené nebo nahlášené s určitým zpožděním. My v této části nebudeme započítávat zpoždění v nahlášení daných událostí.

V celé této kapitole uvažujeme definici 1.5.3. Příslušným počtem a celkovou výší škod, které nastaly v roce C_i a jež byly vyplaceny v čase C_{ij} , $j \in \mathbb{N}_0$, budeme označovat N_{ij} a S_{ij} . Index j nám vyjadřuje počet let, které uplynuly od vzniku pojistné události. S využitím procesu (1.7) může být počet plateb za škody z roku i zapsán ve formě

$$\begin{aligned} N_{ij} &= \int_{(0, \infty) \times E_M} \sum_{l=1}^k I_{\{t \in C_i, t + (y_1 + \dots + y_l) \in C_{ij}\}} N_M(dt, d(y_r), d(x_r), dk) \\ &= \sum_{n \geq 1} \sum_{l=1}^{K_n} I_{\{T_n \in C_i, T_n + (Y_{n1} + \dots + Y_{nl}) \in C_{ij}\}} \\ &= \sum_{n \geq 1} \sum_{l=1}^{K_n} I_{\{T_n \in C_i, T_{nl} \in C_{ij}\}}, \quad i \in \mathbb{N}, \quad j \in \mathbb{N}_0, \end{aligned}$$

a odpovídajícím způsobem úhrn plateb

$$\begin{aligned}
S_{ij} &= \int_{(0,\infty) \times E_M} \sum_{l=1}^k x_l I_{\{t \in C_i, t+(y_1+\dots+y_l) \in C_{ij}\}} N_M(dt, d(y_r), d(x_r), dk) \\
&= \sum_{n \geq 1} \sum_{l=1}^{K_n} X_{nl} I_{\{T_n \in C_i, T_n+(Y_{n1}+\dots+Y_{nl}) \in C_{ij}\}} \\
&= \sum_{n \geq 1} \sum_{l=1}^{K_n} X_{nl} I_{\{T_n \in C_i, T_{nl} \in C_{ij}\}}, \quad i \in \mathbb{N}, j \in \mathbb{N}_0.
\end{aligned}$$

Naše data zapíšeme pomocí trojúhelníku dvojic (N_{ij}, S_{ij}) , $i = 1, \dots, m, 1 \leq i + j \leq m$:

$$\begin{array}{ccccccc}
(N_{10}, S_{10}) & (N_{11}, S_{11}) & (N_{12}, S_{12}) & \dots & (N_{1,m-1}, S_{1,m-1}) \\
(N_{20}, S_{20}) & (N_{21}, S_{21}) & \dots & (N_{2,m-2}, S_{2,m-2}) & & & \\
\dots & \dots & & & & & \\
(N_{m0}, S_{m0}) & & & & & &
\end{array}$$

Jestliže uvažujeme C_m jako současný rok, pak chain ladder obsahuje úplnou agregovanou roční informaci o škodních platbách v předcházejících letech a současném roce. Tato informace se skládá z počtu plateb N_{ij} a odpovídající úhrnné částky S_{ij} pro škody, které nastaly v roce C_i pro nějaké $i \leq m$ a které byly vypořádány v období C_{ij} pro $i + j \leq m$.

2.2 Mackův model

Praktický problém je následující: Jak můžeme použít informace ve formě minulých a současných pozorování (N_{ij}, S_{ij}) , $i = 1, \dots, m, 1 \leq i + j \leq m$, abychom předpověděli budoucí úhrn škodních plateb S_{im} a počet plateb N_{im} , $i = 1, \dots, m$?

Na tuto otázku se pokouší odpovědět Mackův model, který přidává některé dodatečné předpoklady.

Definice 2.2.1. (Mackův model)

Uvažujme obecný shlukový model z definice 1.5.3 a předpokládejme navíc následující dvě podmínky:

- (M1) *Počty a celkové částky plateb pro škody vzniklé v různých letech jsou nezávislé a stejně rozdělené, tj. procesy $(N_{ij}, S_{ij})_{j=0,1,\dots}$, $i = 1, 2, \dots$, jsou nezávislé a stejně rozdělené.*
- (M2) *Náhodné veličiny N_{ij}, S_{ij} , $i = 1, 2, \dots, j = 0, 1, \dots$ jsou skoro jistě kladné a mají konečnou střední hodnotu. Existují nezáporná, reálná čísla f_j a g_j taková, že pro $i = 1, 2, \dots, j = 0, 1, \dots$, platí:*

$$\mathbb{E}(N_{i,j+1} | N_{i0}, \dots, N_{ij}) = g_j N_{ij}, \quad (2.1)$$

$$\mathbb{E}(S_{i,j+1} | S_{i0}, \dots, S_{ij}) = f_j S_{ij}. \quad (2.2)$$

Pro zjednodušení zápisu použijeme standardní posloupnosti

$$(N_j)_{j \geq 0} = (N_{1,j})_{j \geq 0} \quad \text{a} \quad (S_j)_{j \geq 0} = (S_{1,j})_{j \geq 0}.$$

Pak podmínky (2.1) a (2.2) můžeme zapsat následujícím způsobem:

$$\mathbb{E}(N_{j+1}|N_0, \dots, N_j) = g_j N_j, \quad (2.3)$$

$$\mathbb{E}(S_{j+1}|S_0, \dots, S_j) = f_j S_j. \quad (2.4)$$

Požadavek konečnosti momentů $\mathbb{E}N_j$ a $\mathbb{E}S_j$ v podmínce (M2) je potřeba, abychom měli jistotu, že podmíněné střední hodnoty (2.3) a (2.4) jsou konečné s pravděpodobností 1. Podmínka (M2) je v literatuře často modifikována. Například je běžné používat méně restriktivní podmínky

$$\mathbb{E}(N_{j+1}|N_j) = g_j N_j \quad \text{a} \quad \mathbb{E}(S_{j+1}|S_j) = f_j S_j, \quad j \geq 0.$$

Nyní se zaměříme pouze na celkovou výši plateb S_{ij} a pokusíme se odvodit budoucí odhad pro celkovou výši plateb v následujících letech a základní vlastnosti tohoto odhadu. Teorie pro počet plateb N_{ij} by se odvíjela analogicky.

V této části budeme postupovat podle knihy [6], ale na rozdíl od ní my budeme uvažovat úhrny škod a ne počty škod.

2.2.1 Asymptotické vlastnosti metody chain ladder

V Mackově modelu jsou pro budoucí odhady škod klíčové odhady konstant f_j , kterým se obvykle říká vývojové faktory. Odhad těchto faktorů je založen na standardní asymptotické teorii, kterou nyní probereme. Konkrétně odhady f_j vznikají pomocí momentové metody odhadů.

Předpoklad (M1) nám říká, že historie škodních událostí, které nastanou v různých letech C_i , mají stejné rozdělení a jsou vzájemně nezávislé. Tyto vlastnosti nám umožňují použití standardní asymptotické teorie pro $m \rightarrow \infty$. Z podmínky (M1) se posloupnost $(S_{ij})_{i \geq 1}$ skládá z nezávislých, stejně rozdělených náhodných veličin pro každé pevné $j \geq 0$. V m -tém roce, řekněme současnosti, jsou známy pouze minulé a současné úhrny plateb. Z Glivenkova-Cantelliho lemmatu plyne pro pevné $j \geq 0$:

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{m-j} \sum_{i=1}^{m-j} I_{[0,x]}(S_{ij}) - P(S_j \leq x) \right| \xrightarrow{s.j.} 0, \quad m \rightarrow \infty.$$

Podobně pro pevné $j \geq 0$ silný zákon velkých čísel implikuje silnou konzistenci pro výběrové odhady $\mathbb{E}S_j$, $\text{var } S_j$ a $\text{cov}(S_{j_1}, S_{j_2})$, jakmile $m \rightarrow \infty$:

$$\bar{S}_m^{(j)} = \frac{1}{m-j} \sum_{i=1}^{m-j} S_{ij} \xrightarrow{s.j.} \mathbb{E}S_j, \quad (2.5)$$

$$[s_m^{(j)}]^2 = \frac{1}{m-j-1} \sum_{i=1}^{m-j} (S_{ij} - \bar{S}_m^{(j)})^2 \xrightarrow{s.j.} \text{var } S_j, \quad (2.6)$$

$$\begin{aligned} \bar{\gamma}_m(j_1, j_2) &= \frac{1}{m-j_2} \sum_{i=1}^{m-j_2} (S_{i,j_1} - \bar{S}_m^{(j_1)}) (S_{i,j_2} - \bar{S}_m^{(j_2)}) \\ &\xrightarrow{s.j.} \text{cov}(S_{j_1}, S_{j_2}), \quad j_1 < j_2. \end{aligned} \quad (2.7)$$

Abychom mohli použít silný zákon velkých čísel, předpokládáme ve vzorci (2.5) konečnou střední hodnotu $\mathbb{E}S_j < \infty$ a ve vzorcích (2.6), (2.7) konečný rozptyl $\text{var} S_j < \infty$. Kdybychom chtěli odvodit asymptotický interval spolehlivosti pro odhady, aplikovali bychom mnohorozměrnou centrální limitní větu na nezávislé, stejně rozdělené náhodné vektory. Ale my se momentálně zaměřujeme na získání odhadů parametrů.

Díky (2.4) pozorujeme, že pro $j = 0, 1, \dots$ platí

$$\mathbb{E}S_{j+1} = \mathbb{E}[\mathbb{E}(S_{j+1}|S_0, \dots, S_j)] = f_j \mathbb{E}S_j.$$

Protože předpokládáme, že S_j je kladné pro každé j , tak také máme $\mathbb{E}S_j > 0$, proto

$$f_j = \frac{\mathbb{E}S_{j+1}}{\mathbb{E}S_j}, \quad j = 0, 1, \dots \quad (2.8)$$

Tento vztah a silný zákon velkých čísel pro $\bar{S}_m^{(j)}$ v (2.5) doporučuje nahradit střední hodnoty v (2.8) odpovídajícími průměry, což dává odhad pro f_j momentovou metodou:

$$\hat{f}_j^{(m)} = \frac{\bar{S}_m^{(j+1)}}{\bar{S}_{m-1}^{(j)}} = \frac{\sum_{i=1}^{m-j-1} S_{i,j+1}}{\sum_{i=1}^{m-j-1} S_{ij}}, \quad j = 0, \dots, m-2. \quad (2.9)$$

Tento odhad f_j se nazývá chain ladder odhad.

Věta 2.2.2. (*Asymptotické vlastnosti chain ladder odhadu*)

Předpokládejme podmínky Mackova modelu.

1. Odhad $\hat{f}_j^{(m)}$ metodou chain ladder je silně konzistentní odhad svého deterministického protějšku f_j . To znamená, že pro pevné $j = 0, 1, \dots$ platí:

$$\hat{f}_j^{(m)} \xrightarrow{s.j.} f_j \quad \text{pro } m \rightarrow \infty.$$

2. Pokud dodáme předpoklad $\text{var} S_j < \infty$ pro $j = 0, 1, \dots$, pak $\hat{f}_j^{(m)}$ je asymptoticky normální, tj. splňuje centrální limitní větu

$$\sqrt{m} (\hat{f}_j^{(m)} - f_j) \xrightarrow{d} N \left(0, (\mathbb{E}S_j)^{-2} [\mathbb{E}S_{j+1}^2 - f_j^2 \mathbb{E}S_j^2] \right).$$

Důkaz. 1. Konzistence plyne přímo ze silného zákona velkých čísel (2.5) a vztahu (2.8):

$$\hat{f}_j^{(m)} = \frac{\bar{S}_m^{(j+1)}}{\bar{S}_{m-1}^{(j)}} \xrightarrow{s.j.} \frac{\mathbb{E}S_{j+1}}{\mathbb{E}S_j} = f_j.$$

2. Nejprve pozorujeme, že

$$\begin{aligned} \sqrt{m} (\hat{f}_j^{(m)} - f_j) &= \sqrt{m} \left(\frac{\bar{S}_m^{(j+1)}}{\bar{S}_{m-1}^{(j)}} - f_j \right) = \frac{\sqrt{m}}{\bar{S}_{m-1}^{(j)}} (\bar{S}_m^{(j+1)} - f_j \bar{S}_{m-1}^{(j)}) \\ &= \frac{\sqrt{m/(m-j-1)}}{\bar{S}_{m-1}^{(j)}} \frac{1}{\sqrt{m-j-1}} \sum_{i=1}^{m-j-1} (S_{i,j+1} - f_j S_{ij}). \end{aligned}$$

Z podmínky (M1) v Mackově modelu vidíme, že posloupnost $(S_{i,j+1} - f_j S_{ij})_{i \geq 1}$ má nezávislé, stejně rozdělené členy pro pevné j , z (2.2) plyne,

že tato posloupnost má nulovou střední hodnotu a z předpokladu věty má konečný rozptyl. Aplikace centrální limitní věty v kombinaci se silným zákonem velkých čísel $\bar{S}_{m-1}^{(j)} \xrightarrow{s.j.} \mathbb{E}S_j$ implikuje, že

$$\sqrt{m} (\hat{f}_j^{(m)} - f_j) \xrightarrow{d} N(0, (\mathbb{E}S_j)^{-2} \text{var}(S_{j+1} - f_j S_j)). \quad (2.10)$$

Z přímého výpočtu platí:

$$\text{var}(S_{j+1} - f_j S_j) = \text{var} S_{j+1} + f_j^2 \text{var} S_j - 2f_j \text{cov}(S_j, S_{j+1}).$$

Provedeme úpravy některých členů:

$$\begin{aligned} \text{cov}(S_j, S_{j+1}) &= \mathbb{E}S_{j+1}S_j - \mathbb{E}S_{j+1}\mathbb{E}S_j \\ &= \mathbb{E}[S_j\mathbb{E}(S_{j+1}|S_j)] - \mathbb{E}[\mathbb{E}(S_{j+1}|S_j)]\mathbb{E}S_j \\ &= f_j \text{var} S_j, \\ \text{var} S_{j+1} &= \mathbb{E}S_{j+1}^2 - f_j^2(\mathbb{E}S_j)^2. \end{aligned}$$

Proto asymptotický rozptyl v (2.10) je dán vztahem

$$\frac{\text{var}(S_{j+1} - f_j S_j)}{(\mathbb{E}S_j)^2} = \frac{\mathbb{E}S_{j+1}^2 - f_j^2\mathbb{E}S_j^2}{(\mathbb{E}S_j)^2}.$$

Z tohoto výrazu plyne náš důkaz. □

Asymptotický rozptyl v centrální limitní větě pro $\hat{f}_j^{(m)}$ může být aproximován momentovými odhady pro $\mathbb{E}S_j^2$, $\mathbb{E}S_{j+1}^2$ a f_j . To znamená, že asymptotický rozptyl $(\mathbb{E}S_j)^{-2} (\mathbb{E}S_{j+1}^2 - f_j^2\mathbb{E}S_j^2)$ je aproximován silně konzistentním odhadem

$$\frac{\sum_{i=1}^{m-j-1} [S_{i,j+1}^2 - (\hat{f}_j^{(m)})^2 S_{ij}^2]}{(m-j)^{-1} (\sum_{i=1}^{m-j} S_{ij})^2}, \quad j = 0, \dots, m-2.$$

2.2.2 Momenty chain ladder odhadů

Odhad $\hat{f}_j^{(m)}$ není přesně výběrový odhad pro $\mathbb{E}S_{j+1}/\mathbb{E}S_j$. Tím by byl odhad

$$\frac{\bar{S}_m^{(j+1)}}{\bar{S}_m^{(j)}} = \frac{m-j}{m-j-1} \frac{\sum_{i=1}^{m-j-1} S_{i,j+1}}{\sum_{i=1}^{m-j} S_{ij}} = \frac{m-j}{m-j-1} \hat{f}_j^{(m)}.$$

Nicméně není těžké vidět, že tento odhad má stejné asymptotické vlastnosti jako $\hat{f}_j^{(m)}$. Jeden z důvodů pro volbu $\hat{f}_j^{(m)}$ je fakt, že je to nestranný odhad deterministického protějšku f_j , což dokážeme v následujícím lemmatu.

V této podkapitole budeme často používat následující σ -algebry:

$$\mathcal{G}_j^{(m)} = \sigma(S_{ik}, 0 \leq k \leq j, 1 \leq i+k \leq m), \quad j = 0, 1, \dots,$$

$$\mathcal{F}_{ij} = \sigma(S_{i0}, \dots, S_{ij}), \quad i = 1, 2, \dots, j = 0, 1, \dots$$

Lemma 2.2.3. *(Střední hodnota a kovariance chain ladder odhadů)*

Za předpokladu podmínek Mackova modelu platí:

1. Chain ladder odhady jsou nestranné:

$$\mathbb{E}\hat{f}_j^{(m)} = f_j, \quad j = 0, \dots, m-2.$$

2. Jestliže budeme navíc předpokládat, že $\text{var } S_j < \infty$ pro $j \in \mathbb{N}_0$, pak chain ladder odhady jsou nekorelované:

$$\text{cov}(\hat{f}_{j_1}^{(m)}, \hat{f}_{j_2}^{(m)}) = 0, \quad 0 \leq j_1 < j_2 \leq m-2.$$

Důkaz. 1. Protože $\sum_{i=1}^{m-j-1} S_{ij}$ je měřitelná vzhledem ke $\mathcal{G}_j^{(m)}$, platí následující vztah

$$\mathbb{E}(\hat{f}_j^{(m)} | \mathcal{G}_j^{(m)}) = \frac{\sum_{i=1}^{m-j-1} \mathbb{E}(S_{i,j+1} | \mathcal{G}_j^{(m)})}{\sum_{i=1}^{m-j-1} S_{ij}}.$$

Na základě podmínky (M1) je náhodná veličina $S_{i,j+1}$ nezávislá na $(S_{lk})_{k>0}$ pro $l \neq i$. Protože $\mathcal{F}_{ij} \subset \mathcal{G}_j^{(m)}$ pro $i = 1, \dots, m-j-1$ a $j = 0, \dots, m-2$, dostáváme

$$\mathbb{E}(S_{i,j+1} | \mathcal{G}_j^{(m)}) = \mathbb{E}(S_{i,j+1} | \mathcal{F}_{ij}) = f_j S_{ij}. \quad (2.11)$$

V posledním kroku jsme použili podmínku (M2). Kombinací předchozích rovností obdržíme

$$\mathbb{E}(\hat{f}_j^{(m)} | \mathcal{G}_j^{(m)}) = f_j. \quad (2.12)$$

Aplikací střední hodnoty na obě strany (2.12) dostáváme $\mathbb{E}\hat{f}_j^{(m)} = f_j$, proto je $\hat{f}_j^{(m)}$ nestranný odhad f_j .

2. Nekorelovanost chain ladder odhadů lze opět ukázat pomocí podmiňování. Předpokládejme $j_1 < j_2$. Pak odhad $\hat{f}_{j_1}^{(m)}$ je měřitelný vzhledem k $\mathcal{G}_{j_2}^{(m)}$, a tudíž použitím (2.12) platí

$$\mathbb{E}(\hat{f}_{j_1}^{(m)} \hat{f}_{j_2}^{(m)} | \mathcal{G}_{j_2}^{(m)}) = \hat{f}_{j_1}^{(m)} \mathbb{E}(\hat{f}_{j_2}^{(m)} | \mathcal{G}_{j_2}^{(m)}) = \hat{f}_{j_1}^{(m)} f_{j_2}.$$

Aplikováním střední hodnoty vidíme, že

$$\begin{aligned} \text{cov}(\hat{f}_{j_1}^{(m)}, \hat{f}_{j_2}^{(m)}) &= \mathbb{E}\hat{f}_{j_1}^{(m)} \hat{f}_{j_2}^{(m)} - \mathbb{E}\hat{f}_{j_1}^{(m)} \mathbb{E}\hat{f}_{j_2}^{(m)} \\ &= \mathbb{E}\hat{f}_{j_1}^{(m)} f_{j_2} - f_{j_1} f_{j_2} = 0. \end{aligned}$$

□

Rozptyl chain ladder odhadu $\hat{f}_j^{(m)}$ není tak lehce zjistitelný, protože se jedná o podíl dvou závislých náhodných veličin.

Pro následující výpočty zavedeme značení:

$$\tilde{S}_{ij} = S_{i,j+1} - f_j S_{ij}, \quad i = 1, 2, \dots, j = 0, 1, \dots$$

Předpokládáme, že $\text{var } S_j < \infty$ pro $j = 0, 1, \dots$. Použitím (2.11) platí

$$\mathbb{E}(\tilde{S}_{ij} | \mathcal{G}_j^{(m)}) = \mathbb{E}(\tilde{S}_{ij} | \mathcal{F}_{ij}) = 0, \quad (2.13)$$

což spolu s (2.12) dává

$$\begin{aligned}\text{var}(\hat{f}_j^{(m)}|\mathcal{G}_j^{(m)}) &= \mathbb{E} \left[\left(\frac{\sum_{i=1}^{m-j-1} \tilde{S}_{ij}}{\sum_{i=1}^{m-j-1} S_{ij}} \right)^2 \mid \mathcal{G}_j^{(m)} \right] = \frac{1}{\sum_{i=1}^{m-j-1} S_{ij}} \text{var} \left(\sum_{i=1}^{m-j-1} \tilde{S}_{ij} \mid \mathcal{G}_j^{(m)} \right) \\ &= \frac{\sum_{i=1}^{m-j-1} \text{var}(\tilde{S}_{ij}|\mathcal{G}_j^{(m)})}{\left(\sum_{i=1}^{m-j-1} S_{ij}\right)^2} + \frac{\sum_{1 \leq i \neq k \leq m-j-1} \text{cov}(\tilde{S}_{ij}, \tilde{S}_{kj}|\mathcal{G}_j^{(m)})}{\left(\sum_{i=1}^{m-j-1} S_{ij}\right)^2}.\end{aligned}$$

Pozorujeme, že

$$\begin{aligned}\text{var}(\tilde{S}_{ij}|\mathcal{G}_j^{(m)}) &= \text{var}(\tilde{S}_{ij}|\mathcal{F}_{ij}), \\ \text{cov}(\tilde{S}_{ij}, \tilde{S}_{kj}|\mathcal{G}_j^{(m)}) &= \mathbb{E}(\tilde{S}_{ij}\tilde{S}_{kj}|\mathcal{F}_{ij} \vee \mathcal{F}_{kj}),\end{aligned}$$

kde $\mathcal{F}_{ij} \vee \mathcal{F}_{kj}$ je nejmenší σ -algebra obsahující σ -algebry \mathcal{F}_{ij} a \mathcal{F}_{kj} , $i, k = 1, \dots, m-j-1$. Kovariance je nulová pro $i < k$:

$$\begin{aligned}\mathbb{E}(\tilde{S}_{ij}\tilde{S}_{kj}|\mathcal{F}_{ij} \vee \mathcal{F}_{kj}) &= \mathbb{E}[\mathbb{E}(\tilde{S}_{ij}\tilde{S}_{kj}|\mathcal{F}_{ij} \vee \mathcal{F}_{k,j+1})|\mathcal{F}_{ij} \vee \mathcal{F}_{kj}] \\ &= \mathbb{E}[\tilde{S}_{kj} \mathbb{E}(\tilde{S}_{ij}|\mathcal{F}_{ij} \vee \mathcal{F}_{k,j+1})|\mathcal{F}_{ij} \vee \mathcal{F}_{kj}] \\ &= \mathbb{E}[\tilde{S}_{kj} \mathbb{E}(\tilde{S}_{ij}|\mathcal{F}_{ij})|\mathcal{F}_{ij} \vee \mathcal{F}_{kj}] = 0,\end{aligned}\tag{2.14}$$

kde jsme použili nezávislost \tilde{S}_{ij} a $\mathcal{F}_{k,j+1}$ a podmínku (M2) Mackova modelu. Čímž jsme dokázali, že

$$\text{var}(\hat{f}_j^{(m)}|\mathcal{G}_j^{(m)}) = \frac{\sum_{i=1}^{m-j-1} \text{var}(\tilde{S}_{ij}|\mathcal{F}_{ij})}{\left(\sum_{i=1}^{m-j-1} S_{ij}\right)^2}, \quad j = 0, \dots, m-2.\tag{2.15}$$

V tomto bodě si musíme uvědomit, že podmínky Mackova modelu jsou nepostačující při vyhodnocování podmíněných rozptylů. Tato myšlenka není překvapující, protože podmínka (M2) vypovídá pouze o podmíněné střední hodnotě S_{ij} , nikoliv o druhých momentech. Proto k podmínkám (M1) a (M2) Mackova modelu přidáme dodatečnou podmínku:

(M3) Předpokládejme, že $\text{var} S_j < \infty$, $j = 0, 1, \dots$, a existují nezáporná reálná čísla σ_j^2 taková, že

$$\text{var}(S_{j+1}|S_0, \dots, S_j) = \sigma_j^2 S_j, \quad j = 0, 1, \dots\tag{2.16}$$

Za platnosti podmínky (M2) lze (2.16) vyjádřit také jako

$$\mathbb{E}[(S_{j+1} - f_j S_j)^2 | S_0, \dots, S_j] = \sigma_j^2 S_j, \quad j = 0, 1, \dots\tag{2.17}$$

Za předpokladu podmínky (M3) jsme nyní schopni určit $\text{var} \hat{f}_j^{(m)}$.

Lemma 2.2.4. (*Rozptyl chain ladder odhadu $\hat{f}_j^{(m)}$*)

Za podmínek Mackova modelu, včetně přidané podmínky (M3), máme pro každé $j = 0, \dots, m-2$:

$$\text{var} \hat{f}_j^{(m)} = \sigma_j^2 \mathbb{E} \left(\sum_{i=1}^{m-j-1} S_{ij} \right)^{-1}.\tag{2.18}$$

Důkaz. Z předpokladu (2.16) a vztahu (2.13) přímo dostáváme

$$\text{var} \left(\tilde{S}_{ij} | \mathcal{G}_j^{(m)} \right) = \text{var} \left(\tilde{S}_{ij} | \mathcal{F}_{ij} \right) = \sigma_j^2 S_{ij}. \quad (2.19)$$

Pak z (2.15) získáme

$$\text{var} \left(\hat{f}_j^{(m)} | \mathcal{G}_j^{(m)} \right) = \frac{\sum_{i=1}^{m-j-1} \sigma_j^2 S_{ij}}{\left(\sum_{i=1}^{m-j-1} S_{ij} \right)^2} = \frac{\sigma_j^2}{\sum_{i=1}^{m-j-1} S_{ij}}. \quad (2.20)$$

Nepodmíněný rozptyl $\hat{f}_j^{(m)}$ daný (2.18) dostaneme aplikací vztahu

$$\begin{aligned} \text{var} \hat{f}_j^{(m)} &= \mathbb{E} \left[\text{var} \left(\hat{f}_j^{(m)} | \mathcal{G}_j^{(m)} \right) \right] + \text{var} \left[\mathbb{E} \left(\hat{f}_j^{(m)} | \mathcal{G}_j^{(m)} \right) \right] \\ &= \mathbb{E} \left[\text{var} \left(\hat{f}_j^{(m)} | \mathcal{G}_j^{(m)} \right) \right], \end{aligned}$$

kde využíváme informace, že $\mathbb{E} \left(\hat{f}_j^{(m)} | \mathcal{G}_j^{(m)} \right) = f_j$, viz (2.12). \square

Odhady hodnot σ_j^2 mohou být odvozené pomocí momentové metody. Pokud použijeme střední hodnotu na rovnici (2.17) definující σ_j^2 , obdržíme

$$\mathbb{E} (S_{j+1} - f_j S_j)^2 = \sigma_j^2 \mathbb{E} S_j.$$

Nyní nahradíme střední hodnotu na obou stranách výběrovými odhady:

$$\frac{1}{m-j-1} \sum_{i=1}^{m-j-1} (S_{i,j+1} - f_j S_{ij})^2 = \hat{\sigma}_j^2 \frac{1}{m-j-1} \sum_{i=1}^{m-j-1} S_{ij}.$$

Nakonec nahrazením f_j odhadem $\hat{f}_j^{(m)}$ dostáváme

$$\left[\hat{\sigma}_j^{(m)} \right]^2 = \frac{\sum_{i=1}^{m-j-1} (S_{i,j+1} - \hat{f}_j^{(m)} S_{ij})^2}{\sum_{i=1}^{m-j-1} S_{ij}}, \quad j = 0, \dots, m-2. \quad (2.21)$$

Můžeme si všimnout, že tento odhad je pro $j = m-2$ roven 0. Odhad (2.21) je záporně vychýlený. Vychýlení je dáno v následující větě.

Věta 2.2.5. (Střední hodnota odhadu $\left[\hat{\sigma}_j^{(m)} \right]^2$)

Za předpokladů Mackova modelu a podmínky (M3) platí pro odhad $\left[\hat{\sigma}_j^{(m)} \right]^2$ parametru σ_j^2 , daný vztahem (2.21), následující:

$$\mathbb{E} \left[\hat{\sigma}_j^{(m)} \right]^2 = \sigma_j^2 - \sigma_j^2 \mathbb{E} \frac{\sum_{i=1}^{m-j-1} S_{ij}^2}{\left(\sum_{i=1}^{m-j-1} S_{ij} \right)^2} \quad j = 0, \dots, m-2.$$

Důkaz. Nejdříve odhad $\left[\hat{\sigma}_j^{(m)} \right]^2$ rozepíšeme s využitím vztahu $\tilde{S}_{ij} = S_{i,j+1} - f_j S_{ij}$:

$$\begin{aligned} \left[\hat{\sigma}_j^{(m)} \right]^2 &= \frac{\sum_{i=1}^{m-j-1} (S_{i,j+1} - \hat{f}_j^{(m)} S_{ij})^2}{\sum_{i=1}^{m-j-1} S_{ij}} = \frac{\sum_{i=1}^{m-j-1} (\tilde{S}_{ij} - (\hat{f}_j^{(m)} - f_j) S_{ij})^2}{\sum_{i=1}^{m-j-1} S_{ij}} \\ &= \frac{1}{\sum_{i=1}^{m-j-1} S_{ij}} \sum_{i=1}^{m-j-1} \left(\tilde{S}_{ij}^2 - 2 (\hat{f}_j^{(m)} - f_j) S_{ij} \tilde{S}_{ij} + (\hat{f}_j^{(m)} - f_j)^2 S_{ij}^2 \right). \end{aligned}$$

Odvodíme podmíněnou střední hodnotu odhadu:

$$\begin{aligned} \mathbb{E} \left(\left[\hat{\sigma}_j^{(m)} \right]^2 \mid \mathcal{G}_j^{(m)} \right) &= \frac{1}{\sum_{i=1}^{m-j-1} S_{ij}} \sum_{i=1}^{m-j-1} \left[\mathbb{E} \left(\tilde{S}_{ij}^2 \mid \mathcal{G}_j^{(m)} \right) \right. \\ &\quad \left. - 2S_{ij} \mathbb{E} \left(\left(\hat{f}_j^{(m)} - f_j \right) \tilde{S}_{ij} \mid \mathcal{G}_j^{(m)} \right) + S_{ij}^2 \mathbb{E} \left(\left(\hat{f}_j^{(m)} - f_j \right)^2 \mid \mathcal{G}_j^{(m)} \right) \right]. \end{aligned} \quad (2.22)$$

Nyní upravíme prostřední člen pomocí (2.9):

$$\begin{aligned} \mathbb{E} \left(\left(\hat{f}_j^{(m)} - f_j \right) \tilde{S}_{ij} \mid \mathcal{G}_j^{(m)} \right) &= \mathbb{E} \left(\frac{\sum_{k=1}^{m-j-1} \tilde{S}_{kj}}{\sum_{k=1}^{m-j-1} S_{kj}} \tilde{S}_{ij} \mid \mathcal{G}_j^{(m)} \right) \\ &= \frac{1}{\sum_{k=1}^{m-j-1} S_{kj}} \mathbb{E} \left(\sum_{k=1}^{m-j-1} \tilde{S}_{kj} \tilde{S}_{ij} \mid \mathcal{G}_j^{(m)} \right). \end{aligned}$$

Jak již víme z (2.14), podmíněná střední hodnota $\mathbb{E} \left(\tilde{S}_{kj} \tilde{S}_{ij} \mid \mathcal{G}_j^{(m)} \right) = \mathbb{E} \left(\tilde{S}_{kj} \tilde{S}_{ij} \mid \mathcal{F}_{ij} \vee \mathcal{F}_{kj} \right)$ je nenulová pouze pro případ $i = k$:

$$\mathbb{E} \left(\tilde{S}_{ij}^2 \mid \mathcal{G}_j^{(m)} \right) = S_{ij} \sigma_j^2,$$

viz (2.19) a (2.13). Použijeme-li ve (2.22) na třetí člen (2.12) a (2.20), dostáváme:

$$\begin{aligned} \mathbb{E} \left(\left[\hat{\sigma}_j^{(m)} \right]^2 \mid \mathcal{G}_j^{(m)} \right) &= \frac{1}{\sum_{i=1}^{m-j-1} S_{ij}} \sum_{i=1}^{m-j-1} \left(S_{ij} \sigma_j^2 - \frac{2S_{ij}^2 \sigma_j^2}{\sum_{k=1}^{m-j-1} S_{kj}} + \frac{S_{ij}^2 \sigma_j^2}{\sum_{k=1}^{m-j-1} S_{kj}} \right) \\ &= \sigma_j^2 - \sigma_j^2 \frac{\sum_{i=1}^{m-j-1} S_{ij}^2}{\left(\sum_{i=1}^{m-j-1} S_{ij} \right)^2}. \end{aligned}$$

Výsledný vztah dostaneme použitím střední hodnoty:

$$\mathbb{E} \left[\hat{\sigma}_j^{(m)} \right]^2 = \sigma_j^2 - \sigma_j^2 \mathbb{E} \frac{\sum_{i=1}^{m-j-1} S_{ij}^2}{\left(\sum_{i=1}^{m-j-1} S_{ij} \right)^2}.$$

□

V článku [3] byl navržen jiný způsob odhadu parametru σ_j^2 :

$$\left[\tilde{\sigma}_j^{(m)} \right]^2 = \frac{1}{m-j-2} \sum_{i=1}^{m-j-1} S_{ij} \left(\frac{S_{i,j+1}}{S_{ij}} - \hat{f}_j^{(m)} \right)^2, \quad j = 0, \dots, m-3. \quad (2.23)$$

Takovýto odhad je nestranný odhad σ_j^2 , $j = 0, \dots, m-3$, což dokážeme v následující větě. Její důkaz je proveden podle [4].

Věta 2.2.6. (*Nestrannost odhadu $\left[\tilde{\sigma}_j^{(m)} \right]^2$*)

Za předpokladů Mackova modelu a podmínky (M3) je odhad $\left[\tilde{\sigma}_j^{(m)} \right]^2$, $j = 0, \dots, m-3$, daný vzorcem (2.23), nestranným odhadem σ_j^2 .

Důkaz. Vzorec (2.23) může být přepsán následujícím způsobem:

$$\begin{aligned} (m-j-2) [\tilde{\sigma}_j^{(m)}]^2 &= \sum_{i=1}^{m-j-1} \left(\frac{S_{i,j+1}^2}{S_{ij}} - 2S_{i,j+1}\hat{f}_j^{(m)} + S_{ij} [\hat{f}_j^{(m)}]^2 \right) \\ &= \sum_{i=1}^{m-j-1} \frac{S_{i,j+1}^2}{S_{ij}} - \sum_{i=1}^{m-j-1} S_{ij} [\hat{f}_j^{(m)}]^2, \end{aligned} \quad (2.24)$$

použili jsme vztah $\sum_{i=1}^{m-j-1} S_{i,j+1} = \hat{f}_j^{(m)} \sum_{i=1}^{m-j-1} S_{ij}$ podle definice $\hat{f}_j^{(m)}$. Aplikujeme podmíněnou střední hodnotu na výraz (2.24) a dostaneme:

$$\begin{aligned} \mathbb{E} \left((m-j-2) [\tilde{\sigma}_j^{(m)}]^2 \mid \mathcal{G}_j^{(m)} \right) &= \sum_{i=1}^{m-j-1} \frac{\mathbb{E}(S_{i,j+1}^2 \mid \mathcal{G}_j^{(m)})}{S_{ij}} \\ &\quad - \sum_{i=1}^{m-j-1} S_{ij} \mathbb{E}([\hat{f}_j^{(m)}]^2 \mid \mathcal{G}_j^{(m)}). \end{aligned} \quad (2.25)$$

Díky nezávislosti z podmínky (M1) získáváme

$$\begin{aligned} \mathbb{E}(S_{i,j+1}^2 \mid \mathcal{G}_j^{(m)}) &= \mathbb{E}(S_{i,j+1}^2 \mid \mathcal{F}_{ij}) \\ &= \text{var}(S_{i,j+1} \mid \mathcal{F}_{ij}) + (\mathbb{E}(S_{i,j+1} \mid \mathcal{F}_{ij}))^2 \\ &= S_{ij}\sigma_j^2 + (S_{ij}f_j)^2, \end{aligned} \quad (2.26)$$

kde jsme použili vztahů (2.2) a (2.16). Z (2.20) a (2.12) získáme

$$\begin{aligned} \mathbb{E} \left([\hat{f}_j^{(m)}]^2 \mid \mathcal{G}_j^{(m)} \right) &= \text{var}(\hat{f}_j^{(m)} \mid \mathcal{G}_j^{(m)}) + (\mathbb{E}(\hat{f}_j^{(m)} \mid \mathcal{G}_j^{(m)}))^2 \\ &= \frac{\sigma_j^2}{\sum_{i=1}^{m-j-1} S_{ij}} + f_j^2. \end{aligned} \quad (2.27)$$

Vložíme-li (2.26) a (2.27) do (2.25), získáme

$$\begin{aligned} \mathbb{E} \left((m-j-2) [\tilde{\sigma}_j^{(m)}]^2 \mid \mathcal{G}_j^{(m)} \right) &= \sum_{i=1}^{m-j-1} (\sigma_j^2 + S_{ij}f_j^2) \\ &\quad - \sum_{i=1}^{m-j-1} \left(\frac{S_{ij}\sigma_j^2}{\sum_{i=1}^{m-j-1} S_{ij}} + S_{ij}f_j^2 \right) \\ &= (m-j-1)\sigma_j^2 - \sigma_j^2 \\ &= (m-j-2)\sigma_j^2. \end{aligned}$$

Z posledního odvození okamžitě obdržíme $\mathbb{E}([\tilde{\sigma}_j^{(m)}]^2 \mid \mathcal{G}_j^{(m)}) = \sigma_j^2$. Nakonec dostáváme závěr:

$$\mathbb{E} [\tilde{\sigma}_j^{(m)}]^2 = \mathbb{E} \left(\mathbb{E}([\tilde{\sigma}_j^{(m)}]^2 \mid \mathcal{G}_j^{(m)}) \right) = \mathbb{E}\sigma_j^2 = \sigma_j^2.$$

□

Odhad $[\tilde{\sigma}_j^{(m)}]^2$ není definován pro $j = m-2$. Podle [3] můžeme doplnit obvykle exponenciálně klesající řadu $\tilde{\sigma}_0^{(m)}, \dots, \tilde{\sigma}_{m-4}^{(m)}, \tilde{\sigma}_{m-3}^{(m)}$ o jeden člen, například

pomocí logaritmicke-lineární regrese, nebo jednodušeji můžeme požadovat platnost rovnosti:

$$\frac{\tilde{\sigma}_{m-4}^{(m)}}{\tilde{\sigma}_{m-3}^{(m)}} = \frac{\tilde{\sigma}_{m-3}^{(m)}}{\tilde{\sigma}_{m-2}^{(m)}},$$

alespoň pokud $\tilde{\sigma}_{m-4}^{(m)} > \tilde{\sigma}_{m-3}^{(m)}$. Tato poslední možnost vede k odhadu

$$\left[\tilde{\sigma}_{m-2}^{(m)}\right]^2 = \min \left(\frac{\left[\tilde{\sigma}_{m-3}^{(m)}\right]^4}{\left[\tilde{\sigma}_{m-4}^{(m)}\right]^2}, \min \left(\left[\tilde{\sigma}_{m-4}^{(m)}\right]^2, \left[\tilde{\sigma}_{m-3}^{(m)}\right]^2 \right) \right).$$

2.2.3 Predikce

V této části se budeme věnovat predikci budoucích úhrnů plateb. Připomeňme, že když chceme předpovědět veličinu S_{j+k} , $k \in \mathbb{N}$, na základě S_0, \dots, S_j , tak nejlepší předpověď minimalizuje střední kvadratickou chybu $\mathbb{E}(S_{j+k} - M)^2$ přes třídu všech náhodných veličin M , které mají konečný rozptyl a jsou měřitelné vzhledem k σ -algebře \mathcal{F}_j generované náhodnými veličinami S_0, \dots, S_j , a řešením je podmíněná střední hodnota $\mathbb{E}[S_{j+k}|S_0, \dots, S_j]$. Střední kvadratickou chybu predikce \hat{S}_{j+k} náhodné veličiny S_{j+k} budeme značit

$$\text{mse } \hat{S}_{j+k} = \mathbb{E} \left(\hat{S}_{j+k} - S_{j+k} \right)^2.$$

Budeme také používat podmíněnou střední kvadratickou chybu, což je náhodná veličina

$$\text{mse}_{\mathcal{F}_j} \hat{S}_{j+k} = \mathbb{E} \left(\left(\hat{S}_{j+k} - S_{j+k} \right)^2 \middle| \mathcal{F}_j \right).$$

Předpokládejme podmínky Mackova modelu a navíc podmínku (M3). Představujme si m jako současný čas a budeme chtít předpovědět budoucí úhrn plateb. Nejdříve se zaměříme na předpověď o jeden krok, tj. chceme předpovědět $S_{i,m-i+1}$, $i = 1, \dots, m$. Z podmínky (M2) máme, že nejlepší predikce je $\mathbb{E}(S_{i,m-i+1} | \mathcal{F}_{i,m-i}) = f_{m-i} S_{i,m-i}$. Protože hodnoty f_{m-i} neznáme, nahradíme je jejich odhady $\hat{f}_{m-i}^{(m)}$. Tím získáme následující:

$$\hat{S}_{i,m-i+1} = \hat{f}_{m-i}^{(m)} S_{i,m-i}, \quad i = 2, \dots, m. \quad (2.28)$$

Tato úprava však vede k tomu, že $\hat{S}_{i,m-i+1}$ již neminimalizuje střední kvadratickou chybu $\mathbb{E}(S_{i,m-i+1} - M)^2$ přes třídu všech náhodných veličin M , které mají konečný rozptyl a jsou měřitelné vzhledem k σ -algebře $\mathcal{F}_{i,m-i}$. Když se podíváme na podmínku (M3), rozptyl nejlepší predikce je dán hodnotou

$$\begin{aligned} \mathbb{E}(S_{i,m-i+1} - f_{m-i} S_{i,m-i})^2 &= \mathbb{E}[\text{var}(S_{i,m-i+1} | \mathcal{F}_{i,m-i})] \\ &= \mathbb{E}[\sigma_{m-i}^2 S_{i,m-i}^2] \\ &= \sigma_{m-i}^2 \mathbb{E} S_{m-i}. \end{aligned} \quad (2.29)$$

Rozptyl $\sigma_{m-i}^2 \mathbb{E} S_{m-i}$ je minimum pro chybu predikce veličiny $S_{i,m-i+1}$. Pozorujeme, že

$$\begin{aligned} \text{mse } \hat{S}_{i,m-i+1} &= \mathbb{E}(\hat{S}_{i,m-i+1} - S_{i,m-i+1})^2 \\ &= \mathbb{E}(S_{i,m-i+1} - f_{m-i} S_{i,m-i})^2 + \mathbb{E} \left[(\hat{f}_{m-i}^{(m)} - f_{m-i})^2 S_{i,m-i}^2 \right] \\ &= \sigma_{m-i}^2 \mathbb{E} S_{m-i} + \text{var } \hat{f}_{m-i}^{(m)} \mathbb{E} S_{m-i}^2. \end{aligned} \quad (2.30)$$

Použili jsme (2.29) a nezávislost $(S_{i,m-i}, S_{i,m-i+1})$ a $\hat{f}_{m-i}^{(m)}$, kterou si můžeme uvědomit z definice odhadu $\hat{f}_{m-i}^{(m)}$. Vidíme, že tento odhad ovlivňují pouze veličiny $S_{k,m-i+1}, S_{k,m-i}, k = 1, \dots, i-1$, které jsou nezávislé na $(S_{i,m-i}, S_{i,m-i+1})$ podle (M1).

Dohromady s rozptylem $\text{var } \hat{f}_{m-i}^{(m)}$ z lemmatu 2.2.4 dostáváme následující výsledek pro chybu predikce.

Lemma 2.2.7. (*Střední kvadratická chyba jednokrokové predikce*)

Předpokládejme podmínky Mackova modelu a také podmínku (M3). Střední kvadratická chyba jednokrokové predikce $\hat{S}_{i,m-i+1}$ definované v (2.28) je dána:

$$\text{mse } \hat{S}_{i,m-i+1} = \sigma_{m-i}^2 \left[\mathbb{E} S_{m-i} + \mathbb{E} \left(\sum_{k=1}^{i-1} S_{k,m-i} \right)^{-1} \mathbb{E} S_{m-i}^2 \right], \quad i = 2, \dots, m.$$

Důkaz. Lemma plyne z (2.30) a lemmatu 2.2.4. □

Z nezávislosti $\hat{f}_{m-i}^{(m)}$ a $S_{i,m-i}$, lemmatu 2.2.3 a vztahu (2.8) plyne pro střední hodnotu predikce

$$\mathbb{E} \hat{S}_{i,m-i+1} = \mathbb{E} \hat{f}_{m-i}^{(m)} \mathbb{E} S_{i,m-i} = f_{m-i} \mathbb{E} S_{i,m-i} = \mathbb{E} S_{i,m-i+1}.$$

Znamená to, že $\hat{S}_{i,m-i+1}$ je nestranná předpověď.

Podmínka (M2) v Mackově modelu nespécifikuje pouze jednokrokovou predikci S_{j+1} pomocí $\mathbb{E}(S_{j+1}|S_0, \dots, S_j) = f_j S_j$, ale také definuje nejlepší predikci (ve smyslu střední kvadratické chyby) S_{j+k} při daných S_0, \dots, S_j pro každé $k \in \mathbb{N}$. Použitím vlastností podmíněné střední hodnoty a indukce je nejlepší k -kroková predikce S_{j+k} daná výrazem

$$\begin{aligned} \mathbb{E}[S_{j+k}|S_0, \dots, S_j] &= \mathbb{E}[\mathbb{E}(S_{j+k}|S_0, \dots, S_{j+k-1})|S_0, \dots, S_j] \\ &= f_{j+k-1} \mathbb{E}[S_{j+k-1}|S_0, \dots, S_j] \\ &= f_{j+k-1} \cdots f_j S_j. \end{aligned} \tag{2.31}$$

Odpovídající chybu predikce získáme výpočtem

$$\begin{aligned} &\mathbb{E}(S_{j+k} - f_{j+k-1} \cdots f_j S_j)^2 \\ &= \mathbb{E}(S_{j+k} - f_{j+k-1} S_{j+k-1})^2 + f_{j+k-1}^2 \mathbb{E}(S_{j+k-1} - f_{j+k-2} S_{j+k-2})^2 \\ &\quad + \cdots + f_{j+k-1}^2 \cdots f_{j+1}^2 \mathbb{E}(S_{j+1} - f_j S_j)^2. \end{aligned}$$

Z (2.17) dostáváme:

$$\begin{aligned} &\mathbb{E}(S_{j+k} - f_{j+k-1} \cdots f_j S_j)^2 \\ &= \sigma_{j+k-1}^2 \mathbb{E} S_{j+k-1} + f_{j+k-1}^2 \sigma_{j+k-2}^2 \mathbb{E} S_{j+k-2} + \cdots + f_{j+k-1}^2 \cdots f_{j+1}^2 \sigma_j^2 \mathbb{E} S_j \\ &= \mathbb{E} S_j \sum_{l=0}^{k-1} f_j \cdots f_{j+l-1} \sigma_{j+l}^2 [f_{j+l+1} \cdots f_{j+k-1}]^2. \end{aligned}$$

Vezmeme vztah (2.31) jako počáteční bod při hledání aproximace nejlepší předpovědi $S_{i,m-i+k}$ danou minulými a současnými pozorováními $S_{i,0}, \dots, S_{i,m-i}$. Nejlepší předpovědí je náhodná veličina

$$\mathbb{E}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) = f_{m-i+k-1} \cdots f_{m-i} S_{i,m-i}, \quad k \in \mathbb{N}. \tag{2.32}$$

Kandidát na aproximaci se získává nahrazením neznámých parametrů f_l jejich chain ladder odhady:

$$\hat{S}_{i,m-i+k} = \hat{f}_{m-i+k-1}^{(m)} \cdots \hat{f}_{m-i}^{(m)} S_{i,m-i}. \quad (2.33)$$

Jelikož odhady $\hat{f}_j^{(m)}$, viz (2.9), máme pouze pro $j \leq m-2$, můžeme $\hat{S}_{i,m-i+k}$ stanovit jen pro $k \geq i-1$, tj. vývojový trojúhelník dat doplníme na čtverec:

$$\begin{array}{cccccc} S_{10} & S_{11} & S_{12} & \cdots & S_{1,m-1} \\ S_{20} & S_{21} & \cdots & S_{2,m-2} & \hat{S}_{2,m-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_{m0} & \hat{S}_{m1} & \cdots & \cdots & \hat{S}_{m,m-1} \end{array}$$

Jak již bylo vysvětleno v kontextu jedнокrokové predikce, $\hat{S}_{i,m-i+k}$ nemá vlastnosti nejlepší predikce $S_{i,m-i+k}$ (ve smyslu střední kvadratické chyby) dané minulými a současnými pozorováními S_{ij} , $i+j \leq m$.

Díky nekorelovanosti chain ladder odhadů vývojových faktorů (lemma 2.2.3) si opět můžeme povšimnout, že $\hat{S}_{i,m-i+k}$ je nestranná předpověď $S_{i,m-i+k}$:

$$\mathbb{E} \hat{S}_{i,m-i+k} = f_{m-i+k-1}^{(m)} \cdots f_{m-i}^{(m)} \mathbb{E} S_{i,m-i} = \mathbb{E} S_{i,m-i+k}. \quad (2.34)$$

Střední kvadratická chyba je vyjádřena v následujících dvou lemmatech.

Lemma 2.2.8. (Podmíněná střední kvadratická chyba k -krokové predikce)

Za předpokladů Mackova modelu s přidáním podmínkou (M3) platí:

$$\text{mse}_{\mathcal{F}_{i,m-i}} \hat{S}_{i,m-i+k} = \text{var}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) + \left(\mathbb{E}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) - \hat{S}_{i,m-i+k} \right)^2, \quad (2.35)$$

kde

$$\text{var}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) = S_{i,m-i} \sum_{j=m-i}^{m-i+k-1} f_{m-i} \cdots f_{j-1} \sigma_j^2 f_{j+1}^2 \cdots f_{m-i+k-1}^2,$$

$$\begin{aligned} & \left(\mathbb{E}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) - \hat{S}_{i,m-i+k} \right)^2 \\ &= S_{i,m-i}^2 \left(f_{m-i} \cdots f_{m-i+k-1} - \hat{f}_{m-i}^{(m)} \cdots \hat{f}_{m-i+k-1}^{(m)} \right)^2. \end{aligned}$$

Důkaz. Použijeme postup podle [3]. Obecně platí $\mathbb{E}(X - a)^2 = \text{var} X + (\mathbb{E} X - a)^2$ a z toho plyne (2.35). Na důkaz zbytku věty použijeme zkrácené značení:

$$\mathbb{E}_i(X) = \mathbb{E}(X | \mathcal{F}_{i,m-i})$$

$$\text{var}_i(X) = \text{var}(X | \mathcal{F}_{i,m-i}).$$

Opakovaným použitím předpokladů (M2) a (M3) dostaneme

$$\begin{aligned} \text{var}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) &= \text{var}_i(S_{i,m-i+k}) \\ &= \mathbb{E}_i(\text{var}(S_{i,m-i+k} | \mathcal{F}_{i,m-i+k-1})) + \text{var}_i(\mathbb{E}(S_{i,m-i+k} | \mathcal{F}_{i,m-i+k-1})) \\ &= \mathbb{E}_i(S_{i,m-i+k-1}) \sigma_{m-i+k-1}^2 + \text{var}_i(S_{i,m-i+k-1}) f_{m-i+k-1}^2 \\ &= \mathbb{E}_i(S_{i,m-i+k-2}) f_{m-i+k-2} \sigma_{m-i+k-1}^2 + \mathbb{E}_i(S_{i,m-i+k-2}) \sigma_{m-i+k-2}^2 f_{m-i+k-1}^2 \\ &\quad + \text{var}_i(S_{i,m-i+k-2}) f_{m-i+k-2}^2 f_{m-i+k-1}^2 \\ &= \dots \\ &= S_{i,m-i} \sum_{j=m-i}^{m-i+k-1} f_{m-i} \cdots f_{j-1} \sigma_j^2 f_{j+1}^2 \cdots f_{m-i+k-1}^2, \end{aligned} \quad (2.36)$$

protože $\text{var}_i(S_{i,m-i}) = 0$.

Díky (2.32) obdržíme druhý sčítanec (2.35):

$$\begin{aligned} & \left(\mathbb{E}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) - \hat{S}_{i,m-i+k} \right)^2 \\ &= S_{i,m-i}^2 \left(f_{m-i} \cdots f_{m-i+k-1} - \hat{f}_{m-i}^{(m)} \cdots \hat{f}_{m-i+k-1}^{(m)} \right)^2. \end{aligned} \quad (2.37)$$

□

Lemma 2.2.9. (*Střední kvadratická chyba k-krokové predikce*)

Předpokládejme podmínky Mackova modelu a navíc podmínku (M3). Střední kvadratická chyba mse $\hat{S}_{i,m-i+k}$ k-krokové predikce $\hat{S}_{i,m-i+k}$ definované v (2.33) je dána:

$$\begin{aligned} \text{mse } \hat{S}_{i,m-i+k} &= \mathbb{E} \left(S_{i,m-i+k} - \hat{S}_{i,m-i+k} \right)^2 \\ &= \mathbb{E} \left(S_{i,m-i+k} - \mathbb{E}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) \right)^2 + \mathbb{E} \left(\mathbb{E}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) - \hat{S}_{i,m-i+k} \right)^2, \end{aligned}$$

kde

$$\begin{aligned} & \mathbb{E} \left(S_{i,m-i+k} - \mathbb{E}(S_{i,m-i+k} | \mathcal{F}_{i,m-i}) \right)^2 \\ &= \mathbb{E} S_{m-i} \sum_{j=m-i}^{m-i+k-1} [f_{m-i+k-1} \cdots f_{j+1}]^2 \sigma_j^2 f_{j-1} \cdots f_{m-i}. \end{aligned}$$

Důkaz. Aplikujeme střední hodnotu na lemma 2.2.8. □

2.2.4 Výpočet predikce rezervy

V této podkapitole budeme vycházet z článku [3]. Vypočítáme predikci rezervy na pojistná plnění, její střední kvadratickou chybu a odhad této chyby. Pro tento výpočet nám bude stačit doplnit náš trojúhelník dat na čtverec pomocí předpovědí budoucích úhrnů plateb.

Rezerva na pojistná plnění je dána vztahem:

$$R_i = S_{i,m-1} - S_{i,m-i}, \quad i = 2, \dots, m.$$

Predikce R_i proto bude $\hat{R}_i = \hat{S}_{i,m-1} - S_{i,m-i}$. Využijeme vztahu (2.33) a dostáváme

$$\hat{R}_i = S_{i,m-i} \left(\hat{f}_{m-2}^{(m)} \cdots \hat{f}_{m-i}^{(m)} - 1 \right), \quad i = 2, \dots, m.$$

Protože $\hat{S}_{i,m-1}$ je podle (2.34) nestranná předpověď $S_{i,m-1}$, je také \hat{R}_i nestrannou predikcí R_i .

Nyní odvodíme podmíněnou střední kvadratickou chybu predikce \hat{R}_i :

$$\begin{aligned} \text{mse}_{\mathcal{F}_{i,m-i}} \hat{R}_i &= \mathbb{E} \left(\left(\hat{R}_i - R_i \right)^2 | \mathcal{F}_{i,m-i} \right) \\ &= \mathbb{E} \left(\left(\hat{S}_{i,m-1} - S_{i,m-1} \right)^2 | \mathcal{F}_{i,m-i} \right) \\ &= \text{mse}_{\mathcal{F}_{i,m-i}} \hat{S}_{i,m-1}. \end{aligned}$$

Později budeme potřebovat odhad σ_j^2 , $j = 0, \dots, m-3$. Použijeme odhad vyjádřený vzorcem (2.23), o kterém jsme si již dokázali, že je nestranný. Protože

m je pevné, budeme ve značení vynechávat horní index (m) a psát pouze $\tilde{\sigma}_j^2$. Podobně budeme psát pouze \hat{f}_j pro odhad f_j .

Z lemmatu 2.2.8 pro $k = i - 1$ vidíme, že platí:

$$\text{mse}_{\mathcal{F}_{i,m-i}} \hat{S}_{i,m-1} = \text{var}(S_{i,m-1} | \mathcal{F}_{i,m-i}) + \left(\mathbb{E}(S_{i,m-1} | \mathcal{F}_{i,m-i}) - \hat{S}_{i,m-1} \right)^2.$$

A tedy stačí najít odhady těchto dvou sčítanců, které vyjadřují $\text{mse}_{\mathcal{F}_{i,m-i}} \hat{R}_i$. Ve výrazu (2.36) nahradíme parametry f_j a σ_j^2 jejich odhady \hat{f}_j a $\tilde{\sigma}_j^2$, tedy odhadneme $\text{var}(S_{i,m-1} | \mathcal{F}_{i,m-i})$ následujícím výrazem:

$$S_{i,m-i} \sum_{j=m-i}^{m-2} \hat{f}_{m-i} \cdots \hat{f}_{j-1} \tilde{\sigma}_j^2 \hat{f}_{j+1}^2 \cdots \hat{f}_{m-2}^2 = \hat{S}_{i,m-1}^2 \sum_{j=m-i}^{m-2} \frac{\tilde{\sigma}_j^2 / \hat{f}_j^2}{\hat{S}_{ij}}.$$

Ve výrazu (2.37) nemůžeme jednoduše nahradit f_j pomocí \hat{f}_j , protože tím bychom dostali 0. Proto zde použijeme jinou aproximaci. Napíšeme-li, že

$$\begin{aligned} F &= f_{m-i} \cdots f_{m-2} - \hat{f}_{m-i} \cdots \hat{f}_{m-2} \\ &= M_{m-i} + \dots + M_{m-2}, \end{aligned}$$

kde

$$M_j = \hat{f}_{m-i} \cdots \hat{f}_{j-1} (f_j - \hat{f}_j) f_{j+1} \cdots f_{m-2}, \quad j = m-i, \dots, m-2,$$

získáme

$$\begin{aligned} F^2 &= (M_{m-i} + \dots + M_{m-2})^2 \\ &= \sum_{j=m-i}^{m-2} M_j^2 + 2 \sum_{k < j} M_j M_k. \end{aligned}$$

Nyní M_j^2 aproximuje $\mathbb{E}(M_j^2 | \mathcal{G}_j^{(m)})$ a $M_j M_k$, $k < j$ nahradíme $\mathbb{E}(M_j M_k | \mathcal{G}_j^{(m)})$. Protože $\mathbb{E}(f_j - \hat{f}_j | \mathcal{G}_j^{(m)}) = 0$, viz (2.12), dostáváme $\mathbb{E}(M_j M_k | \mathcal{G}_j^{(m)}) = 0$ pro $k < j$. Ze vztahu (2.20) dostaneme

$$\mathbb{E}(M_j^2 | \mathcal{G}_j^{(m)}) = \frac{\hat{f}_{m-i}^2 \cdots \hat{f}_{j-1}^2 \tilde{\sigma}_j^2 \hat{f}_{j+1}^2 \cdots \hat{f}_{m-2}^2}{\sum_{i=1}^{m-j-1} S_{ij}}.$$

Vzmemme-li to dohromady, výraz $F^2 = \left(\sum_{j=m-i}^{m-2} M_j \right)^2$ nahradíme součtem $\sum_{j=m-i}^{m-2} \mathbb{E}(M_j^2 | \mathcal{G}_j^{(m)})$ a protože všechny výrazy tohoto součtu jsou kladné, můžeme nahradit všechny neznámé parametry f_j , σ_j^2 jejich nestrannými odhady \hat{f}_j , $\tilde{\sigma}_j^2$. Dohromady odhadujeme $F^2 = \left(f_{m-i} \cdots f_{m-2} - \hat{f}_{m-i} \cdots \hat{f}_{m-2} \right)^2$ výrazem

$$\sum_{j=m-i}^{m-2} \frac{\hat{f}_{m-i}^2 \cdots \hat{f}_{j-1}^2 \tilde{\sigma}_j^2 \hat{f}_{j+1}^2 \cdots \hat{f}_{m-2}^2}{\sum_{i=1}^{m-j-1} S_{ij}} = \hat{f}_{m-i}^2 \cdots \hat{f}_{m-2}^2 \sum_{j=m-i}^{m-2} \frac{\tilde{\sigma}_j^2 / \hat{f}_j^2}{\sum_{i=1}^{m-j-1} S_{ij}}.$$

Výsledný odhad podmíněné střední kvadratické chyby $\text{mse}_{\mathcal{F}_{i,m-i}} \hat{R}_i$ vypadá takto:

$$\widehat{\text{mse}_{\mathcal{F}_{i,m-i}} \hat{R}_i} = \hat{S}_{i,m-1}^2 \sum_{j=m-i}^{m-2} \frac{\tilde{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{S}_{ij}} + \frac{1}{\sum_{k=1}^{m-j-1} S_{kj}} \right).$$

2.2.5 Příklad

Teorii této kapitoly jsme aplikovali na reálných datech, které jsme převzali z materiálu [1], jejich hodnoty jsou uvedeny v příloze v tabulce 4.9. Jedná se o data společnosti TrygVesta, konkrétně o pojištění odpovědnosti za škodu z provozu motorového vozidla. Tabulka 4.9 znázorňuje úhrny plateb za období 18 let. Všechny výpočty jsme provedli s použitím programu Microsoft Excel (2007).

Nejdříve jsme vypočítali odhady \hat{f}_j , $\tilde{\sigma}_j^2$ a $\hat{\sigma}_j^2$, jejichž výsledky jsou uvedeny v tabulce 2.1.

j	\hat{f}_j	$\tilde{\sigma}_j^2$	$\hat{\sigma}_j^2$
0	3,2154	2,4074	1,4842
1	1,9616	2,8292	2,8138
2	1,6633	1,0669	0,9119
3	1,3927	0,1923	0,1804
4	1,2379	0,2392	0,2138
5	1,1497	0,2534	0,2555
6	1,0848	0,1088	0,1146
7	1,0649	0,0883	0,0870
8	1,0325	0,0263	0,0163
9	1,0353	0,0372	0,0401
10	1,0223	0,0280	0,0271
11	1,0134	0,0081	0,0081
12	1,0224	0,0017	0,0012
13	1,0080	0,0058	0,0042
14	0,9996	0,0003	0,0002
15	1,0020	0,0004	0,0002
16	1,0001	0,0003	0,0000

Tabulka 2.1: Odhady \hat{f}_j , $\tilde{\sigma}_j^2$ a $\hat{\sigma}_j^2$.

Na základě odhadů \hat{f}_j jsme spočítali predikce S_{ij} pro $i+j > 18$ a doplnili jsme takto tabulku 4.9 na čtverec. Výsledná matice je zobrazena v příloze v tabulce 4.10. Na základě těchto odhadů jsme mohli provést výpočty predikce rezervy pro úhrny plateb a odhad podmíněných středních kvadratických chyb predikce rezervy. Výsledky jsou zobrazeny v tabulce 2.2 této podpodkapitoly.

i	\hat{R}_i	$\widehat{\text{mse}}_{\mathcal{F}_{i,m-i}} \hat{R}_i$
2	0,0053	0,0528
3	0,2029	0,1548
4	0,1993	0,2626
5	0,9678	0,9618
6	4,2606	1,7168
7	6,3012	3,3011
8	10,4925	9,5366
9	13,1820	13,7746
10	20,4710	22,4722
11	24,9110	33,4478
12	30,0616	45,7596
13	63,3530	133,3319
14	66,0828	143,1580
15	114,4673	246,6801
16	162,1867	676,5269
17	167,4072	1897,3606
18	169,9607	2966,5880

Tabulka 2.2: Predikce \hat{R}_i rezervy a odhady $\widehat{\text{mse}}_{\mathcal{F}_{i,m-i}} \hat{R}_i$ podmíněné střední kvadratické chyby predikce rezervy pro jednotlivé roky.

3. Poissonův shlukový model

V této kapitole budeme vycházet z článku [2].

V předchozí kapitole jsme uvažovali S_{ij} a N_{ij} jako úhrny plateb a úhrny počtu plateb za škody z roku i , které byly vypořádány v časovém období C_{ij} . V této kapitole se pro zjednodušení omezme pouze na jeden rok, předpokládejme $i = 1$. Pro $j \in \mathbb{N}_0$ pak budeme označovat N'_j počty plateb, které byly uhrazeny se zpožděním j a S'_j úhrny plateb, které byly uhrazeny se zpožděním j , pro které vzhledem k předchozímu modelu platí následující:

$$N_j = N'_0 + \dots + N'_j,$$

$$S_j = S'_0 + \dots + S'_j.$$

Pokud bychom vytvořili z hodnot N'_j a S'_j trojúhelník, získali bychom nekumulativní vývojový trojúhelník pro výše plateb a počty plateb.

3.1 Popis modelu

Uvažujeme stochastický model pro počty plateb N'_j a odpovídající výše plateb S'_j , $j \in \mathbb{N}_0$. Model je daný následujícími podmínkami, které budeme předpokládat po celou tuto kapitolu.

Nechť M je počet událostí, které nastanou v daném roce, s rozdělením daným pravděpodobnostmi

$$q_m = P(M = m), \quad m \in \mathbb{N}_0.$$

Nechť m -tá událost způsobuje proud K_m plateb od pojistitele k pojištěnému v průběhu následujících let. Předpokládáme, že k -tá z těchto plateb je vypořádána v roce Y_{mk} . Dále předpokládáme, že (K_m) je posloupnost nezávislých, stejně rozdělených náhodných veličin s Poissonovým rozdělením s parametrem μ a že $(Y_{mk})_{m,k \in \mathbb{N}}$ představuje rodinu nezávislých, stejně rozdělených náhodných veličin se společným rozdělením

$$p_j = P(Y_{11} = j), \quad j \in \mathbb{N}_0.$$

Nakonec předpokládejme, že M , (K_m) a (Y_{mk}) jsou nezávislé. Píšeme

$$N'_j = \sum_{m=1}^M \sum_{k=1}^{K_m} I_{\{Y_{mk}=j\}}, \quad j \in \mathbb{N}_0,$$

tj. N'_j je počet plateb za škody nastalé v daném roce a vypořádané se zpožděním j let.

Na tento model se také můžeme dívat tak, že vznikl zobecněním Neymanova-Scottové Poissonova procesu (definice 1.3.4) s tím, že rodičovský proces není Poissonův, ale nabývá M událostí v čase $j = 0$.

Předpokládáme dále, že $(X_{mk})_{m,k \in \mathbb{N}}$ je rodina nezávislých, stejně rozdělených nezáporných náhodných veličin, nezávislých na M , (K_m) a (Y_{mk}) . Interpretujeme X_{mk} jako k -tou platbu m -té události. Pak

$$S'_j = \sum_{m=1}^M \sum_{k=1}^{K_m} X_{mk} I_{\{Y_{mk}=j\}}, \quad j \in \mathbb{N}_0,$$

je celkový úhrn plateb za škody nastalé v daném roce a vypořádané se zpožděním j let. Znamená to, že příslušný shlukový bodový proces jsme pomocí nezávislého kótování (definice 1.4.2) obohatili o kóty představující výšky úhrad.

V celé této kapitole budeme používat Laplaceovu-Stieltjesovu transformaci náhodné veličiny M a budeme ji značit následujícím způsobem:

$$L(\gamma) = \mathbb{E} e^{-\gamma M} = \sum_{m=0}^{\infty} q_m e^{-\gamma m}, \quad \gamma \geq 0.$$

Její l -tá derivace je

$$L^{(l)}(\gamma) = (-1)^l \mathbb{E} M^l e^{-\gamma M}, \quad \gamma > 0, \quad l \in \mathbb{N}_0.$$

Navíc definujeme

$$R_l(\gamma) = -\frac{L^{(l+1)}(\gamma)}{L^{(l)}(\gamma)} = \frac{\mathbb{E} M^{l+1} e^{-\gamma M}}{\mathbb{E} M^l e^{-\gamma M}}, \quad \gamma > 0, \quad l \in \mathbb{N}_0.$$

Nakonec značíme

$$\theta_j = \mu \sum_{i=0}^j p_i, \quad j \in \mathbb{N}_0;$$

připomeňme, že μ je intenzita Poissonova počtu plateb k dané události a p_i je pravděpodobnost, že platba bude provedena v i -tém roce po uskutečnění události.

3.2 Predikce

Budeme se snažit předpovědět budoucí počet plateb N'_{j+l+1} a odpovídající úhrn plateb S'_{j+l+1} , $l \in \mathbb{N}_0$ na základě předchozího pozorování počtů plateb N'_0, \dots, N'_j . Tedy určíme predikci \hat{S}'_{j+l+1} a \hat{N}'_{j+l+1} jako $\hat{N}'_{j+l+1} = \mathbb{E}(N'_{j+l+1} | \mathcal{F}_j)$ a $\hat{S}'_{j+l+1} = \mathbb{E}(S'_{j+l+1} | \mathcal{F}_j)$, kde $\mathcal{F}_j = \sigma(N'_0, \dots, N'_j) = \sigma(N_0, \dots, N_j)$, $j \in \mathbb{N}_0$.

Nezávislost (X_{mk}) a zbylých náhodných veličin v modelu implikuje

$$\hat{S}'_{j+l+1} = \mathbb{E} X_{11} \mathbb{E}(N'_{j+l+1} | \mathcal{F}_j) = \mathbb{E} X_{11} \hat{N}'_{j+l+1}, \quad l \in \mathbb{N}_0. \quad (3.1)$$

Proto se problém predikce S'_{j+l+1} redukuje na predikci N'_{j+l+1} . Naopak, jestliže $X_{mk} = 1$ s.j. pro všechny m, k , pak $N'_j = S'_j$, $j \in \mathbb{N}_0$. Proto je postačující studovat predikci S'_{j+l+1} na základě N'_0, \dots, N'_j .

Naším cílem je odvodit predikci (3.1) a určit její chybu. Začneme jednkrokovou predikcí, tj. $l = 0$.

3.2.1 Jednkroková predikce

Nyní zformulujeme hlavní výsledek pro predikci S'_{j+1} . Připomeňme, že odpovídající výsledek pro N'_{j+1} plyne položením $X_{mk} = 1$ s.j. pro všechny m, k .

Věta 3.2.1. *(Jednkroková predikce a její chyba)*

Předpokládejme, že $\mathbb{E} M < \infty$ a $\mathbb{E} X_{11} < \infty$. Necht $j \in \mathbb{N}_0$.

1. Predikce \hat{S}'_{j+1} veličiny S'_{j+1} na základě N'_0, \dots, N'_j má tvar

$$\hat{S}'_{j+1} = \mathbb{E} (S'_{j+1} | \mathcal{F}_j) = \mu p_{j+1} \mathbb{E} X_{11} R_{N_j}(\theta_j), \quad (3.2)$$

kde $N_j = N'_0 + \dots + N'_j$.

2. Předpokládejme navíc, že $\text{var } M < \infty$ a $\text{var } X_{11} < \infty$. Pak nepodmíněná chyba predikce pro S'_{j+1} je daná výrazem

$$\begin{aligned} \text{mse } \hat{S}'_{j+1} &= \mathbb{E} \left(S'_{j+1} - \hat{S}'_{j+1} \right)^2 \\ &= \mathbb{E} X_{11}^2 \mu p_{j+1} \mathbb{E} M + (\mathbb{E} X_{11} \mu p_{j+1})^2 \mathbb{E} M^2 - \mathbb{E} (\hat{S}'_{j+1})^2. \end{aligned} \quad (3.3)$$

3. Předpokládejme navíc, že $\text{var } M < \infty$ a $\text{var } X_{11} < \infty$. Pak podmíněná chyba predikce pro S'_{j+1} na základě předchozích pozorování N'_0, \dots, N'_j je

$$\begin{aligned} \text{mse}_{\mathcal{F}_j} \hat{S}'_{j+1} &= \text{var} \left(S'_{j+1} | \mathcal{F}_j \right) = \mathbb{E} X_{11}^2 \mu p_{j+1} R_{N_j}(\theta_j) \\ &\quad + (\mathbb{E} X_{11} \mu p_{j+1})^2 R_{N_j}(\theta_j) \left[R_{N_j+1}(\theta_j) - R_{N_j}(\theta_j) \right], \end{aligned} \quad (3.4)$$

kde $N_j = N'_0 + \dots + N'_j$.

Důkaz. 1. Pro $j \in \mathbb{N}_0$ a $l \in \mathbb{N}$ označme $Z_{jl} = \sum_{k=1}^{K_l} I_{\{Y_{lk}=j\}}$ počet plateb v l -tém proudu plateb, které jsou vypořádány v roce j . Z vlastností Poissonova procesu plyne, že (Z_{jl}) tvoří dvojitou řadu nezávislých náhodných veličin takových, že Z_{jl} má Poissonovo rozdělení s parametrem μp_j . Protože $N'_j = \sum_{l=1}^M Z_{jl}$, podmíněně na M je $(N'_j)_{j \in \mathbb{N}_0}$ posloupnost nezávislých náhodných veličin, které mají Poissonovo rozdělení. Proto

$$\mathbb{E} \left(N'_{j+1} | N'_0, \dots, N'_j, M \right) = M \mu p_{j+1}, \quad j \in \mathbb{N}_0,$$

a

$$\hat{S}'_{j+1} = \mathbb{E} X_{11} \mu p_{j+1} \mathbb{E} (M | \mathcal{F}_j), \quad j \in \mathbb{N}_0. \quad (3.5)$$

Přesněji pro každé $m, j \in \mathbb{N}_0$ a $n_0, \dots, n_j \in \mathbb{N}_0$,

$$\begin{aligned} \mathbb{P}(N'_0 = n_0, \dots, N'_j = n_j, M = m) &= q_m \mathbb{P} \left(\sum_{l=1}^m Z_{0l} = n_0 \right) \cdots \mathbb{P} \left(\sum_{l=1}^m Z_{jl} = n_j \right) \\ &= q_m \prod_{i=0}^j \frac{(m \mu p_i)^{n_i}}{n_i!} e^{-m \mu p_i} \\ &= q_m e^{-m \theta_j} m^{\sum_{i=0}^j n_i} \prod_{i=0}^j \frac{(\mu p_i)^{n_i}}{n_i!}. \end{aligned} \quad (3.6)$$

Usuzujeme, že pro $j \in \mathbb{N}_0$,

$$\mathbb{P}(M = m | N'_0 = n_0, \dots, N'_j = n_j) = \frac{q_m e^{-m \theta_j} m^{\sum_{i=0}^j n_i}}{\sum_{r=0}^{\infty} q_r e^{-r \theta_j} r^{\sum_{i=0}^j n_i}}, \quad (3.7)$$

$m, n_0, \dots, n_j = 0, 1, \dots$ Speciálně:

$$\begin{aligned} \mathbb{E} \left(M | N'_0 = n_0, \dots, N'_j = n_j \right) &= \frac{\sum_{m=0}^{\infty} q_m e^{-m \theta_j} m^{\sum_{i=0}^j n_i + 1}}{\sum_{r=0}^{\infty} q_r e^{-r \theta_j} r^{\sum_{i=0}^j n_i}} \\ &= R_{n_0 + \dots + n_j}(\theta_j). \end{aligned} \quad (3.8)$$

Tedy použitím (3.5) platí (3.2).

2. & 3. Začneme výpočtem chyby predikce S'_{j+1} na základě N'_0, \dots, N'_j . Nejprve pozorujeme, že

$$\text{var} \left(S'_{j+1} | N'_0, \dots, N'_j, M \right) = M \mathbb{E} X_{11}^2 \mu p_{j+1}.$$

Zohledníme-li tento vztah, dostáváme

$$\begin{aligned} \mathbb{E} \left[\left(S'_{j+1} \right)^2 | \mathcal{F}_j \right] &= \mathbb{E} \left[\mathbb{E} \left(\left(S'_{j+1} \right)^2 | N'_0, \dots, N'_j, M \right) | \mathcal{F}_j \right] \\ &= \mathbb{E} \left[\text{var} \left(S'_{j+1} | N'_0, \dots, N'_j, M \right) \right. \\ &\quad \left. + \left(\mathbb{E} \left(S'_{j+1} | N'_0, \dots, N'_j, M \right) \right)^2 | \mathcal{F}_j \right] \\ &= \mathbb{E} \left[M \mathbb{E} X_{11}^2 \mu p_{j+1} + \left(M \mathbb{E} X_{11} \mu p_{j+1} \right)^2 | \mathcal{F}_j \right] \\ &= \mathbb{E} X_{11}^2 \mathbb{E} \left(N'_{j+1} | \mathcal{F}_j \right) + \left(\mathbb{E} X_{11} \mu p_{j+1} \right)^2 \mathbb{E} \left(M^2 | \mathcal{F}_j \right). \end{aligned}$$

Podmíněná chyba predikce tak může být zapsána jako

$$\text{var} \left(S'_{j+1} | \mathcal{F}_j \right) = \mathbb{E} X_{11}^2 \mathbb{E} \left(N'_{j+1} | \mathcal{F}_j \right) + \left(\mathbb{E} X_{11} \mu p_{j+1} \right)^2 \text{var} \left(M | \mathcal{F}_j \right). \quad (3.9)$$

Použitím (3.7) můžeme nahradit podmíněné momenty M odpovídajícími derivacemi L , což vede k (3.4). Spočteme-li střední hodnotu ve vztahu (3.9), obdržíme chybu predikce

$$\begin{aligned} \mathbb{E} \left(S'_{j+1} - \hat{S}'_{j+1} \right)^2 &= \mathbb{E} \left[\text{var} \left(S'_{j+1} | \mathcal{F}_j \right) \right] \\ &= \mathbb{E} X_{11}^2 \mathbb{E} N'_{j+1} + \left(\mathbb{E} X_{11} \mu p_{j+1} \right)^2 \left[\mathbb{E} M^2 - \mathbb{E} \left(\mathbb{E} (M | \mathcal{F}_j) \right)^2 \right] \\ &= \mathbb{E} X_{11}^2 \mu p_{j+1} \mathbb{E} M + \left(\mathbb{E} X_{11} \mu p_{j+1} \right)^2 \mathbb{E} M^2 - \mathbb{E} \left(\hat{S}'_{j+1} \right)^2. \end{aligned}$$

Čímž jsme dokázali 2. a 3. část věty. □

Poznámka 3.2.2. *Jednoduchá horní mez pro nepodmíněnou chybu predikce (3.3) je dána vztahem*

$$\mathbb{E} \left(S'_{j+1} - \hat{S}'_{j+1} \right)^2 \leq \mathbb{E} X_{11}^2 \mu p_{j+1} \mathbb{E} M + \left(\mathbb{E} X_{11} \mu p_{j+1} \right)^2 \mathbb{E} M^2.$$

Vyhodnocení $\mathbb{E} \left(\hat{S}'_{j+1} \right)^2$ v (3.3) je komplikované. Jestliže využijeme posledního důkazu, obdržíme následující explicitní vyjádření:

$$\begin{aligned} \mathbb{E} \left(\hat{S}'_{j+1} \right)^2 &= \left(\mathbb{E} X_{11} \mu p_{j+1} \right)^2 \mathbb{E} \left(\mathbb{E} (M | \mathcal{F}_j) \right)^2 \\ &= \left(\mathbb{E} X_{11} \mu p_{j+1} \right)^2 \mathbb{E} \left(R_{N'_0 + \dots + N'_j}(\theta_j) \right)^2 \\ &= \left(\mathbb{E} X_{11} \mu p_{j+1} \right)^2 \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \mathbb{P}(N'_0 + \dots + N'_j = k, M = m) \left(R_k(\theta_j) \right)^2. \end{aligned}$$

Aplikací (3.6) se dvojitá suma na pravé straně změní na

$$\begin{aligned} &\sum_{k=0}^{\infty} \left[\sum_{m=0}^{\infty} q_m e^{-m\theta_j} m^k \right] \sum_{n_0 + \dots + n_j = k} \prod_{i=0}^j \frac{(\mu p_i)^{n_i}}{n_i!} \left(R_k(\theta_j) \right)^2 \\ &= e^{\theta_j} \sum_{k=0}^{\infty} (-1)^k \frac{L^{(k+1)}(\theta_j)^2}{L^{(k)}(\theta_j)} \mathbb{P}(\Theta_j = k) \\ &= e^{\theta_j} \mathbb{E} \left((-1)^{\Theta_j} \frac{L^{(\Theta_j+1)}(\theta_j)^2}{L^{(\Theta_j)}(\theta_j)} \right), \end{aligned}$$

kde Θ_j má Poissonovo rozdělení s parametrem θ_j .

3.2.2 Vícekroková predikce

V této části se budeme zabývat predikcí o $l + 1$ kroků dopředu. To znamená, že nás zajímají hodnoty \hat{S}'_{j+l+1} , $l \in \mathbb{N}_0$ a odpovídající chyby predikce.

Věta 3.2.3. (*Vícekroková predikce a její chyba*)

Předpokládáme, že $\mathbb{E}M < \infty$, $\mathbb{E}X_{11} < \infty$. Uvažujme $j, l \in \mathbb{N}_0$ pevné.

1. *Predikce \hat{S}'_{j+l+1} veličiny S'_{j+l+1} na základě N'_0, \dots, N'_j má tvar*

$$\hat{S}'_{j+l+1} = \mathbb{E} \left(S'_{j+l+1} | \mathcal{F}_j \right) = \mathbb{E} X_{11} \mu p_{j+l+1} R_{N_j}(\theta_j),$$

kde $N_j = N'_0 + \dots + N'_j$.

2. *Předpokládejme, že $\text{var } M < \infty$ a $\text{var } X_{11} < \infty$. Pak nepodmíněná chyba predikce S'_{j+l+1} je daná následujícím vztahem*

$$\begin{aligned} \text{mse } \hat{S}'_{j+l+1} &= \mathbb{E} \left(S'_{j+l+1} - \hat{S}'_{j+l+1} \right)^2 \\ &= \mathbb{E} X_{11}^2 \mu p_{j+l+1} \mathbb{E} M + (\mathbb{E} X_{11} \mu p_{j+l+1})^2 \mathbb{E} M^2 - \mathbb{E} (\hat{S}'_{j+l+1})^2. \end{aligned}$$

3. *Předpokládejme, že $\text{var } M < \infty$ a $\text{var } X_{11} < \infty$. Pak podmíněná chyba predikce S'_{j+l+1} je daná výrazem*

$$\begin{aligned} \text{mse}_{\mathcal{F}_j} \hat{S}'_{j+l+1} &= \text{var} \left(S'_{j+l+1} | \mathcal{F}_j \right) = \mathbb{E} X_{11}^2 \mu p_{j+l+1} R_{N_j}(\theta_j) \\ &\quad + (\mathbb{E} X_{11} \mu p_{j+l+1})^2 R_{N_j}(\theta_j) \left[R_{N_{j+1}}(\theta_j) - R_{N_j}(\theta_j) \right]. \end{aligned}$$

Důkaz. 1. Začneme pozorováním, že pro $l \in \mathbb{N}_0$,

$$\hat{S}'_{j+l+1} = \mathbb{E} \left[\mathbb{E} \left(S'_{j+l+1} | \mathcal{F}_{j+l} \right) | \mathcal{F}_j \right].$$

A proto, použitím (3.5), obdržíme

$$\begin{aligned} \hat{S}'_{j+l+1} &= \mathbb{E} X_{11} \mu p_{j+l+1} \mathbb{E} \left[\mathbb{E} (M | \mathcal{F}_{j+l}) | \mathcal{F}_j \right] \\ &= \mathbb{E} X_{11} \mu p_{j+l+1} \mathbb{E} (M | \mathcal{F}_j). \end{aligned}$$

Nyní použijeme vztah (3.8) a dokázali jsme bod 1.

2. & 3. Pro podmíněnou chybu predikce S'_{j+l+1} pozorujeme, že pomocí (3.9) platí

$$\begin{aligned} \text{var} \left(S'_{j+l+1} | \mathcal{F}_j \right) &= \mathbb{E} \left(\text{var} \left(S'_{j+l+1} | \mathcal{F}_{j+l} \right) | \mathcal{F}_j \right) + \text{var} \left(\mathbb{E} \left(S'_{j+l+1} | \mathcal{F}_{j+l} \right) | \mathcal{F}_j \right) \\ &= \mathbb{E} X_{11}^2 \hat{N}'_{j+l+1} + (\mathbb{E} X_{11} \mu p_{j+l+1})^2 \mathbb{E} \left(\text{var} (M | \mathcal{F}_{j+l}) | \mathcal{F}_j \right) \\ &\quad + (\mathbb{E} X_{11} \mu p_{j+l+1})^2 \text{var} \left(\mathbb{E} (M | \mathcal{F}_{j+l}) | \mathcal{F}_j \right) \\ &= \mathbb{E} X_{11}^2 \hat{N}'_{j+l+1} + (\mathbb{E} X_{11} \mu p_{j+l+1})^2 \text{var} (M | \mathcal{F}_j). \end{aligned}$$

Podmíněné momenty M můžou být opět vyjádřeny pomocí (3.7). Použijeme-li střední hodnoty, obdržíme nepodmíněnou chybu predikce

$$\begin{aligned}\mathbb{E} \left(S'_{j+l+1} - \hat{S}'_{j+l+1} \right)^2 &= \mathbb{E} \left[\text{var} \left(S'_{j+l+1} | \mathcal{F}_j \right) \right] \\ &= \mathbb{E} X_{11}^2 \mu p_{j+l+1} \mathbb{E} M + (\mathbb{E} X_{11} \mu p_{j+l+1})^2 \left[\mathbb{E} M^2 - \mathbb{E} \left(\mathbb{E} (M | \mathcal{F}_j) \right)^2 \right] \\ &= \mathbb{E} X_{11}^2 \mu p_{j+l+1} \mathbb{E} M + (\mathbb{E} X_{11} \mu p_{j+l+1})^2 \mathbb{E} M^2 - \mathbb{E} \left(\hat{S}'_{j+l+1} \right)^2.\end{aligned}$$

□

Poznámka 3.2.4. *Poznamenejme, že*

$$\hat{S}'_{j+l+1} = \frac{p_{j+l+1}}{p_{j+1}} \hat{S}'_{j+1}, \quad l \in \mathbb{N}_0,$$

za předpokladu, že $p_{j+1} > 0$. Navíc, jestliže $p_{j+l+1} = 0$, pak $\hat{S}'_{j+l+1} = 0$.

3.2.3 Predikce pro (a,b)-rozdělení

V předchozí části jsme předpovídali hodnoty N'_{j+l+1} a S'_{j+l+1} , $l \in \mathbb{N}_0$, na základě N'_0, \dots, N'_j . Při této předpovědi bylo klíčové určení derivací $(-1)^l L^{(l)}(\gamma) = \mathbb{E} M^l e^{-\gamma M}$. V této části předpokládáme, že rozdělení M patří do (a, b) -třídy, viz podmínka (P2) v podpodkapitole 1.5.1, a díky tomu můžeme použít Panjerovu rekurzivní formuli.

Odvodíme rekurzi pro výraz $(-1)^l L^{(l)}(\gamma) = \mathbb{E} M^l e^{-\gamma M}$. Pokud použijeme (a, b) -strukturu a binomický vzorec máme pro $l \in \mathbb{N}$,

$$\begin{aligned}(-1)^l L^{(l)}(\gamma) &= \sum_{n=1}^{\infty} e^{-\gamma n} n^l \left(a + \frac{b}{n} \right) q_{n-1} \\ &= a e^{-\gamma} \sum_{n=0}^{\infty} e^{-\gamma n} (n+1)^l q_n + b e^{-\gamma} \sum_{n=0}^{\infty} e^{-\gamma n} (n+1)^{l-1} q_n \\ &= a e^{-\gamma} \sum_{n=0}^{\infty} e^{-\gamma n} \sum_{r=0}^l \binom{l}{r} n^r q_n + b e^{-\gamma} \sum_{n=0}^{\infty} e^{-\gamma n} \sum_{r=0}^{l-1} \binom{l-1}{r} n^r q_n \\ &= a e^{-\gamma} (-1)^l L^{(l)}(\gamma) + e^{-\gamma} \sum_{r=0}^{l-1} \left[a \binom{l}{r} + b \binom{l-1}{r} \right] (-1)^r L^{(r)}(\gamma).\end{aligned}$$

A proto

$$(-1)^l L^{(l)}(\gamma) = \frac{e^{-\gamma}}{1 - a e^{-\gamma}} \sum_{r=0}^{l-1} \left[a \binom{l}{r} + b \binom{l-1}{r} \right] (-1)^r L^{(r)}(\gamma). \quad (3.10)$$

Nyní uvažujme tři třídy (a, b) -rozdělení.

Tvrzení 3.2.5. *(Rekurzivní vzorce pro derivace Laplaceovy-Stieltjesovy transformace)*

Předpokládejme, že M má rozdělení z (a, b) -třídy.

1. *V Poissonově případě s parametrem b , $\text{Pois}(b)$, $a = 0$, $b > 0$, máme:*

$$\begin{aligned}L(\gamma) &= e^{-b(1-e^{-\gamma})}, \quad \gamma \geq 0, \\ (-1)^l L^{(l)}(\gamma) &= e^{-\gamma} b \sum_{r=0}^{l-1} \binom{l-1}{r} (-1)^r L^{(r)}(\gamma), \quad l \in \mathbb{N}.\end{aligned}$$

2. V binomickém případě, $\text{Bin}(n, p)$, $a = -\frac{p}{1-p}$, $b = \frac{p(n+1)}{1-p}$, $n \in \mathbb{N}$ a $p \in (0, 1)$, máme pro $l \in \mathbb{N}$:

$$L(\gamma) = \left(1 - p(1 - e^{-\gamma})\right)^n, \quad \gamma \geq 0,$$

$$(-1)^l L^{(l)}(\gamma) = \frac{p e^{-\gamma}}{1 - p(1 - e^{-\gamma})} \sum_{r=0}^{l-1} \binom{l}{r} \left[(n+1) \frac{l-r}{l} - 1 \right] (-1)^r L^{(r)}(\gamma).$$

3. V negativně binomickém případě, $a = 1 - p$, $b = (1 - p)(\nu - 1)$, $p \in (0, 1)$ a $\nu > 0$, máme pro $l \in \mathbb{N}$:

$$L(\gamma) = \left(\frac{p}{1 - (1 - p)e^{-\gamma}} \right)^\nu, \quad \gamma \geq 0,$$

$$(-1)^l L^{(l)}(\gamma) = \frac{(1 - p)e^{-\gamma}}{1 - (1 - p)e^{-\gamma}} \sum_{r=0}^{l-1} \binom{l}{r} \left[(\nu - 1) \frac{l-r}{l} + 1 \right] (-1)^r L^{(r)}(\gamma).$$

Důkaz. Důkaz tohoto tvrzení získáme dosazením příslušných hodnot a, b do vztahu (3.10). \square

Poznámka 3.2.6. V Poissonově případě můžeme dostat také jinou rekurzi pro $L^{(k)}$. Definujme polynom H_n stupně n rekurzivně:

$$H_0(x) = 1, \quad H_n(x) = -x \left[H'_{n-1}(x) + H_{n-1}(x) \right], \quad n \in \mathbb{N}, \quad x > 0.$$

Pak výpočtem dostáváme

$$L^{(k)}(\gamma) = H_k(b e^{-\gamma}) L(\gamma), \quad k \in \mathbb{N}_0, \quad \gamma > 0.$$

Speciálně $R_k(\gamma)$ je lomená funkce $b e^{-\gamma}$ pro každé $k \in \mathbb{N}_0$:

$$R_k(\gamma) = -\frac{H_{k+1}(b e^{-\gamma})}{H_k(b e^{-\gamma})} = b e^{-\gamma} \left(1 + \frac{H'_k(b e^{-\gamma})}{H_k(b e^{-\gamma})} \right).$$

Poznámka 3.2.7. Hodnoty $(-1)^l L^{(l)}(\gamma)$ rostou rychle jako funkce l , a proto standardní software začnou dávat hodnotu ∞ i pro mírně velké hodnoty l . Tomuto numerickému problému se můžeme vyhnout, přepíšeme-li (3.10) pomocí výrazů $R_r(\gamma)$, které jsou relevantní pro vzorec predikce předpokládaný v předchozích částech:

$$R_l(\gamma) = \frac{e^{-\gamma}}{1 - a e^{-\gamma}} \sum_{r=0}^l \left[a \binom{l+1}{r} + b \binom{l}{r} \right] (R_r(\gamma) \cdots R_{l-1}(\gamma))^{-1}.$$

Poslední rekurzi pro R_l se vyhneme přímému výpočtu velkých hodnot $|L^{(l)}(\gamma)|$.

3.2.4 Asymptotické chování predikce

Budeme studovat asymptotické chování predikce $\mathbb{E}(S'_{j+1}|N'_0 = n_0, \dots, N'_j = n_j)$, $j \in \mathbb{N}_0$, když pro počet plateb platí $N_j = N'_0 + \dots + N'_j = n_0 + \dots + n_j = k \rightarrow \infty$. Stejná diskuze by se aplikovala na $(l+1)$ -krokovou predikci $\mathbb{E}(S'_{j+l+1}|N'_0 = n_0, \dots, N'_j = n_j)$, $j, l \in \mathbb{N}_0$. S ohledem na výsledek věty 3.2.1 potřebujeme studovat asymptotické chování podílu $R_k(\gamma)$ pro $k \rightarrow \infty$.

Důvodem vedoucím ke zkoumání asymptotického chování $R_k(\gamma)$ pro $k \rightarrow \infty$ je srovnání s Mackovým modelem, který tvrdí, že predikce N'_{j+l+1} na základě N'_0, \dots, N'_j je lineární funkce $k = N_j = N'_0 + \dots + N'_j$. V našem případě je predikce násobkem $R_k(\gamma)$, což nemá žádný důvod být lineární. Nicméně toto pozorování nevylučuje případ, že limita $k^{-1}R_k(\gamma)$ existuje, je konečná a kladná. V takových případech by $R_k(\gamma)$ aproximovalo lineární funkci pro velké k , stejně jako v Mackově modelu. Následující výsledek dává postačující podmínku pro asymptotickou linearitu $R_k(\gamma)$.

Lemma 3.2.8. (*Asymptotická linearita $R_k(\gamma)$*)

Předpokládejme, že $q_n > 0$ pro $n \geq n_0$ a limita

$$\lim_{n \rightarrow \infty} \frac{q_n}{q_n - 1} = e^{-\tau} \in (0, 1] \quad (3.11)$$

existuje. Pak

$$\lim_{k \rightarrow \infty} \frac{R_k(\gamma)}{k} = \frac{1}{\gamma + \tau}.$$

Důkaz. Toto lemma je dokázané v článku [2], strana 225. □

S ohledem na lemma 3.2.8 a větu 3.2.1 docházíme k závěru, že za vhodných podmínek týkajících se rozdělení M , predikce S'_{j+1} na základě N'_0, \dots, N'_j je asymptoticky lineární.

Důsledek 3.2.9. (*Asymptotická linearita predikce*)

Předpokládejme, že $\mathbb{E}X_{11} < \infty$ a že rozdělení M splňuje podmínku (3.11). Pak jakmile $n_0 + \dots + n_j \rightarrow \infty$,

$$\mathbb{E}(S'_{j+1}|N'_0 = n_0, \dots, N'_j = n_j) \sim \frac{\mathbb{E}X_{11}\mu p_{j+1}}{\tau + \theta_j} (n_0 + \dots + n_j).$$

Důkaz. Důkaz plyne z lemmatu 3.2.8 a věty 3.2.1. □

Poznámka 3.2.10. *Pro upřesnění značení: $a_n \sim b_n$, jestliže $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$.*

Nyní se budeme věnovat chování $R_k(\gamma)$ pro rozdělení z (a, b) -třídy.

Negativně binomické rozdělení je jediné rozdělení z (a, b) -třídy splňující podmínku (3.11), kde $e^{-\tau} = 1 - p$.

V případě binomického rozdělení je jasné, že pro $k \rightarrow \infty$ platí:

$$R_k(\gamma) = \frac{\sum_{m=1}^n \binom{m}{n}^k m e^{-\gamma m} q_m}{\sum_{m=1}^n \binom{m}{n}^k e^{-\gamma m} q_m} \rightarrow n.$$

Stejný výsledek dostaneme pro libovolné rozdělení $(q_m)_{m=0, \dots, n}$, $n \in \mathbb{N}$ s $q_n > 0$. V tomto případě, jestliže $\mathbb{E}X_{11} < \infty$, pak pokud $n_0 + \dots + n_j \rightarrow \infty$, platí

$$\mathbb{E}(S'_{j+1}|N'_0 = n_0, \dots, N'_j = n_j) \rightarrow n\mathbb{E}X_{11}\mu p_{j+1}.$$

Nyní se věnujme případu Poissonova rozdělení. Nechť M má Poissonovo rozdělení s parametrem b , pak

$$\mathbb{E}M^k e^{-\gamma M} = e^{-b(1-e^{-\gamma})} \sum_{n=1}^{\infty} n^k \left(\frac{b}{e^\gamma}\right)^n \frac{1}{n!} e^{-be^{-\gamma}}, \quad k \in \mathbb{N}.$$

A proto podíl $R_k(\gamma)$ je roven podílu momentů $\frac{\mathbb{E}(M')^{k+1}}{\mathbb{E}(M')^k}$ pro Poissonovu náhodnou veličinu M' s odlišnou střední hodnotou, například λ . Budeme studovat asymptotické chování těchto podílů pro $k \rightarrow \infty$.

Lemma 3.2.11. (*Asymptotické chování podílů momentů Poissonova rozdělení*)
Nechť M' má Poissonovo rozdělení s parametrem λ . Pak pro $k \rightarrow \infty$ platí

$$\frac{\mathbb{E}(M')^{k+1}}{\mathbb{E}(M')^k} \sim \frac{k}{\log k}.$$

Důkaz. Důkaz tohoto lemmatu nalezneme v článku [2], strana 229. □

Vezmeme-li v potaz poslední lemma 3.2.11 a odstavec před jeho zněním, dojdeme k závěru, že jestliže M má Poissonovo rozdělení s parametrem b a pokud $n_0 + \dots + n_j \rightarrow \infty$, pak

$$\mathbb{E}\left(S'_{j+1} | N'_0 = n_0, \dots, N'_j = n_j\right) \sim \mathbb{E}X_{11} \mu p_{j+1} \frac{n_0 + \dots + n_j}{\log(n_0 + \dots + n_j)}.$$

Na závěr této podkapitoly si řekneme o chování $R_k(\gamma)$ pro velké γ . Když vyhodnocujeme podíl $R_k(\gamma)$ numericky, pozorujeme neobvyklý jev pro velké hodnoty γ : $R_k(\gamma)$ osciluje silněji pro nepřiliš velká k , zatímco tento efekt pozvolna mizí při zvětšujícím se k . Pro malé hodnoty γ toto chování nemůže být pozorováno. Nyní si odvodíme limitní tvrzení vysvětlující tento jev.

Předpokládejme, že

$$\gamma \rightarrow \infty, \quad k = k(\gamma) \quad \text{a} \quad \frac{k}{\gamma} \rightarrow t \in (0, \infty). \quad (3.12)$$

Pro $t > 0$ uvažujeme striktně konvexní funkci

$$h_t(s) = s - t \log s, \quad s > 0,$$

kteřá dosahuje svého minima pro $s = t$. Z tohoto důvodu posloupnost $(h_t(n))_{n \in \mathbb{N}}$ dosahuje svého minima v bodě $n = n_-(t) = \lfloor t \rfloor$, jestliže $n_-(t) \in \mathbb{N}$, nebo v $n = n_+(t) = \lceil t \rceil$ a je možné, že se hodnoty h_t v bodech $n_-(t)$ a $n_+(t)$ shodují. Předpokládáme, že existuje právě jedna hodnota $n = n_*(t)$, ve které h_t nabývá minima. Pro $k \in \mathbb{N}$ máme

$$\mathbb{E}M^k e^{-\gamma M} = q_{n_*(t)} n_*(t)^k e^{-\gamma n_*(t)} + \sum_{n \neq n_*(t)} n^k e^{-\gamma n} q_n.$$

Na základě (3.12) platí

$$q_{n_*(t)} n_*(t)^k e^{-\gamma n_*(t)} = q_{n_*(t)} e^{-\gamma h_{\frac{k}{\gamma}}(n_*(t))} \rightarrow q_{n_*(t)} e^{-\gamma h_t(n_*(t))}. \quad (3.13)$$

Předpoklad, že minimum $h_t(n)$ je jednoznačné pro dané t implikuje, že minimum $h_u(n)$ je také jednoznačné pro u z malého okolí t a dosahuje ho v bodě $n_*(t)$. Proto pro u dostatečně blízko t platí

$$\begin{aligned} \min_{n \neq n_*(t)} h_u(n) &= \min(h_u(n_*(t) - 1), h_u(n_*(t) + 1)) \\ &\rightarrow \min(h_t(n_* - 1), h_t(n_* + 1)), \quad u \rightarrow t. \end{aligned} \quad (3.14)$$

Z tohoto důvodu pro dostatečně velké γ je

$$\begin{aligned} \sum_{n \neq n_*(t)} h_u(n) &= \sum_{n \neq n_*(t)} e^{-\gamma h_{\frac{k}{\gamma}}(n)} q_n \\ &\leq e^{-\gamma \min(h_{k/\gamma}(n_*(t)-1), h_{k/\gamma}(n_*(t)+1))}. \end{aligned} \quad (3.15)$$

Nyní z (3.13)–(3.15) plyne následující

$$\mathbb{E}M^k e^{-\gamma M} \sim q_{n_*(t)} (n_*(t))^k e^{-\gamma n_*(t)},$$

za předpokladu, že $q_{n_*(t)} \neq 0$, a proto za podmínek (3.12) platí

$$R_k(\gamma) \rightarrow n_*(t). \quad (3.16)$$

Pro $j \in \mathbb{N}$ necht

$$a_j = \inf \{t \geq 0 : n_*(t) = j\} = \frac{1}{\log j - \log(j-1)}.$$

Pak $a_j < a_{j+1}$ a ze vztahu (3.16) obdržíme následující výsledek.

Tvrzení 3.2.12. (*Limita $R_k(\gamma)$*)

Pro $j \in \mathbb{N}$ necht $t \in (a_j, a_{j+1})$ je takové, že $q_{n_(t)} \neq 0$. Jestliže k a γ rostou podle (3.12), pak*

$$R_k(\gamma) \rightarrow j.$$

Důkaz. Tvrzení plyne z úvah před jeho formulací. □

Tedy, jestliže k a γ jsou velké a $\frac{k}{\lambda} \in (a_j, a_{j+1})$ pro j , které není příliš velké, pak $R_k(\gamma)$ bude blízko j .

4. Simulace

Pro porovnání metod výpočtů uvedených v předchozích dvou kapitolách jsme provedli dva různé simulační experimenty, které v této kapitole upřesníme a uvedeme výslednou průměrnou relativní absolutní chybu predikce, které jednotlivé metody dosáhly. Veškeré simulace a výpočty byly provedeny v programu Mathematica [8]. Zdrojové kódy programu jsou na příloženém CD.

4.1 Vysvětlení postupu

Data jsme simulovali podle modelu z podkapitoly 3.1. Celkem jsme uvažovali 10 různých let. Pro každý rok $i = 1, \dots, 10$ jsme nezávisle nasimulovali počty M_i a $K_m^{(i)}$ pro $m = 1, \dots, M_i$. Dále jsme vygenerovali realizace posloupností náhodných veličin $(Y_{mk}^{(i)})$ a $(X_{mk}^{(i)})$, kde $m = 1, \dots, M_i$, $k = 1, \dots, K_m^{(i)}$. Z těchto posloupností jsme již snadno dostali úhrny počtů plateb

$$N'_{ij} = \sum_{m=1}^{M_i} \sum_{k=1}^{K_m^{(i)}} I_{\{Y_{mk}^{(i)}=j\}}$$

a úhrny plateb

$$S'_{ij} = \sum_{m=1}^{M_i} \sum_{k=1}^{K_m^{(i)}} X_{mk}^{(i)} I_{\{Y_{mk}^{(i)}=j\}}.$$

Kumulativní počty a úhrny pak jsou $N_{ij} = N'_{i0} + \dots + N'_{ij}$ a $S_{ij} = S'_{i0} + \dots + S'_{ij}$, kde $i = 1, \dots, 10$, $j = 0, \dots, 9$. Konkrétní volby rozdělení a parametrů při těchto simulacích uvedeme v podkapitolách 4.2 a 4.3.

Při této simulaci jsme vlastně vytvořili historii posledních deseti let, na kterou nahlížíme jako na „reálné“ data, která použijeme pro naše výpočty. Současně jsme nasimulovali i „budoucí“ data, se kterými porovnáваме spočtené predikce.

Do výpočtů předpovědí nám vstupovaly pouze data splňující podmínku $i + j \leq 10$. Měli jsme k dispozici dvě metody, a to chain ladder z kapitoly 2 a metodu založenou na Poissonově shlukovém modelu, viz kapitola 3, a na předpokladu, že M má Poissonovo rozdělení s parametrem λ neboli

$$q_m = e^{-\lambda} \frac{\lambda^m}{m!}, \quad m \in \mathbb{N}_0.$$

Oběma metodami se určila predikce \hat{S}_{ij} a \hat{N}_{ij} pro $i + j > 10$. Následně jsme vypočítali relativní absolutní chybu (v procentech) pro $i + j > 10$ podle vztahů:

$$\frac{|S_{ij} - \hat{S}_{ij}|}{S_{ij}} \cdot 100\%, \quad \frac{|N_{ij} - \hat{N}_{ij}|}{N_{ij}} \cdot 100\%.$$

Celá procedura se opakovala 50krát pro každý experiment. Z 50 opakování jsme vypočítali průměrnou relativní absolutní chybu.

U metody chain ladder jsme predikci určili pomocí vztahu (2.33) a u metody založené na Poissonovu shlukovém modelu jsme použili větu 3.2.3. Problém je,

že predikce daná větou 3.2.3 vyžaduje znalost parametrů λ , μ , $\mathbb{E}X_{11}$ a pravděpodobností p_j , $j \in \mathbb{N}_0$. V praxi je musíme odhadnout. K tomu nám nestačí pouze data ve formě kumulativního trojúhelníku, ale využijeme navíc počet událostí, které nastaly v daných letech. Pro odhad parametru λ použijeme průměr počtu pojistných událostí v jednotlivých letech:

$$\hat{\lambda} = \frac{1}{10} \sum_{i=1}^{10} M_i.$$

Tím pádem pravděpodobnosti q_m odhadujeme následovně:

$$\hat{q}_m = e^{-\hat{\lambda}} \frac{(\hat{\lambda})^m}{m!}, \quad m \in \mathbb{N}_0.$$

Parametr μ odhadneme jako průměrný počet plateb na jednu událost v prvním roce:

$$\hat{\mu} = \frac{1}{M_1} \sum_{j=0}^9 N'_{1j}.$$

Ostatní parametry určíme jen z hodnot N'_{ij} a S'_{ij} nekumulativních trojúhelníků. Odhad střední hodnoty náhodné veličiny X_{11} získáme za pomoci vzorce:

$$\widehat{\mathbb{E}X_{11}} = \frac{\sum_{j=0}^9 S'_{1j}}{\sum_{j=0}^9 N'_{1j}}.$$

Pravděpodobnosti p_j odhadneme z následujícího vztahu:

$$\hat{p}_j = \frac{\frac{1}{10-j} \sum_{i=1}^{10-j} N'_{ij}}{\sum_{k=0}^9 N'_{1k}}, \quad j = 0, \dots, 9.$$

4.2 První experiment

Předpokládali jsme, že M_i má rozdělení z (a, b) -třídy, a to konkrétně Poissonovo s parametrem $\lambda = 90$. Rozdělení $K_m^{(i)}$ jsme také volili Poissonovo s parametrem $\mu = 8$. Dalšími simulovanými veličinami byly $(Y_{mk}^{(i)})_{m,k \in \mathbb{N}}$, které jsme generovali z diskrétního rozdělení daného pravděpodobnostmi p_j , které jsme volili lineárně klesající $p_j = 0,15 - j/90$ pro $j = 0, \dots, 9$. Poslední veličiny, které zbývalo nasimulovat, byly $(X_{mk}^{(i)})_{m,k \in \mathbb{N}}$. Tyto veličiny jsme vzali z logaritmicke-normálního rozdělení s parametry 0,01 a 0,5.

Jak již bylo řečeno v podkapitole 4.1, pomocí Mackova modelu a Poissonova shlukového modelu jsme spočítali predikce úhrnů počtů plateb a úhrnů plateb a doplnili tak matici na čtverec. Následně jsme udělali porovnání našich predikcí se simulací celého čtverce a vypočítali jsme relativní absolutní chybu.

Z 50 simulací jsme získali průměrnou relativní absolutní chybu pro úhrny počtů plateb a úhrny plateb oběma metodami. Tabulky 4.1 a 4.2 udávají průměrnou relativní absolutní chybu počtu plateb a tabulky 4.3 a 4.4 udávají průměrnou relativní absolutní chybu úhrnů plateb za škody.

Nejdříve porovnáme průměrnou relativní absolutní chybu úhrnů počtů plateb u obou metod. Porovnáním tabulek 4.1 a 4.2 vidíme, že menší průměrná relativní absolutní chyba nastala u Poissonova shlukového modelu, ale pouze pro hodnoty,

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,80064 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,83006 & 1,28333 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,92707 & 1,42292 & 1,73741 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,16435 & 1,53327 & 1,61697 & 2,00868 \\ 0 & 0 & 0 & 0 & 0 & 1,49188 & 1,98069 & 2,21948 & 2,19031 & 2,50099 \\ 0 & 0 & 0 & 0 & 1,69840 & 2,16586 & 2,67768 & 2,66554 & 3,23862 & 3,52811 \\ 0 & 0 & 0 & 2,18899 & 2,73383 & 3,30844 & 3,54026 & 3,54299 & 3,60150 & 4,01054 \\ 0 & 0 & 2,29637 & 3,85896 & 4,39397 & 4,95881 & 5,10616 & 5,04616 & 5,35398 & 5,39506 \\ 0 & 5,40201 & 6,03398 & 6,58721 & 7,14081 & 7,31259 & 7,38682 & 7,27086 & 7,43746 & 7,71420 \end{pmatrix}$$

Tabulka 4.1: Průměrná relativní absolutní chyba predikce úhrnů počtů plateb v procentech spočítaná pomocí Mackova modelu

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,85697 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,88003 & 1,39139 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,88868 & 1,37627 & 1,72818 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,27319 & 1,80326 & 2,04851 & 2,62063 \\ 0 & 0 & 0 & 0 & 0 & 1,45153 & 1,88386 & 2,08206 & 2,07625 & 2,52429 \\ 0 & 0 & 0 & 0 & 1,67094 & 2,22221 & 2,58440 & 2,46735 & 2,87575 & 3,21593 \\ 0 & 0 & 0 & 2,08612 & 2,68454 & 3,19772 & 3,42033 & 3,49712 & 3,53437 & 3,79466 \\ 0 & 0 & 2,44231 & 3,61650 & 3,98644 & 4,24228 & 4,00685 & 3,93471 & 4,16835 & 4,14074 \\ 0 & 3,96593 & 5,08424 & 5,14348 & 5,54580 & 5,56070 & 5,60036 & 5,60793 & 5,36988 & 5,30239 \end{pmatrix}$$

Tabulka 4.2: Průměrná relativní absolutní chyba predikce úhrnů počtů plateb v procentech spočítaná pomocí Poissonova shlukového modelu

kteří jsou predikováni na základě menšího množství pozorovaných dat (neboli pro větší i) a jedná se o predikci o více kroků. U předpovědi o menší počet kroků, například o jeden krok, dosahujeme podobných výsledků, ale vidíme, že chyba roste s menším množstvím informací vstupujících do výpočtu.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,97924 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,85830 & 1,14956 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,92955 & 1,39684 & 1,65870 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,19196 & 1,75099 & 1,99074 & 2,39038 \\ 0 & 0 & 0 & 0 & 0 & 1,81392 & 2,20751 & 2,47580 & 2,47134 & 2,92286 \\ 0 & 0 & 0 & 0 & 2,05646 & 2,88096 & 3,31527 & 3,29480 & 3,76568 & 4,01835 \\ 0 & 0 & 0 & 2,30852 & 2,81127 & 3,70923 & 4,03798 & 4,08966 & 4,01196 & 4,22652 \\ 0 & 0 & 3,02346 & 4,23952 & 5,04178 & 5,06493 & 5,35627 & 5,41092 & 5,50778 & 5,60242 \\ 0 & 6,62043 & 7,55517 & 8,22618 & 8,66632 & 9,02024 & 8,94893 & 8,87390 & 9,03188 & 9,12949 \end{pmatrix}$$

Tabulka 4.3: Průměrná relativní absolutní chyba predikce úhrnů plateb v procentech spočítaná pomocí Mackova modelu

Při porovnání tabulek 4.3 a 4.4, které znázorňují průměrnou relativní absolutní chybu predikce úhrnů plateb, vidíme, že při předpovědi o menší počet kroků a na základě většího množství dat má Mackův model menší chybu. Zaměříme-li se ale na předpověď o více kroků, pozorujeme, že metoda číslo dva dosahuje menší průměrné relativní absolutní chyby. Navíc je vidět, že v případě predikce o jeden krok chyba roste s menším počtem informací vstupujících do výpočtu.

U obou metod pozorujeme, že průměrná relativní absolutní chyba roste směrem k pravému dolnímu rohu, což je pochopitelné vzhledem k tomu, že se tam

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,98277 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,91431 & 1,46400 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00527 & 1,48307 & 1,88791 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,18514 & 1,89239 & 2,16011 & 2,73162 \\ 0 & 0 & 0 & 0 & 0 & 1,73556 & 2,14663 & 2,59169 & 2,54781 & 2,84353 \\ 0 & 0 & 0 & 0 & 1,99609 & 2,53412 & 2,96371 & 2,80298 & 3,05966 & 3,32347 \\ 0 & 0 & 0 & 2,16987 & 2,83427 & 3,80788 & 3,96442 & 4,09816 & 3,97631 & 4,12431 \\ 0 & 0 & 2,93324 & 4,05272 & 4,45356 & 4,58270 & 4,39149 & 4,36031 & 4,48530 & 4,28851 \\ 0 & 4,61377 & 5,51408 & 5,71453 & 5,80422 & 5,64840 & 5,69957 & 5,72526 & 5,61654 & 5,58409 \end{pmatrix}$$

Tabulka 4.4: Průměrná relativní absolutní chyba predikce úhrnů plateb v procentech spočítaná pomocí Poissonova shlukového modelu

nachází hodnoty, které jsou předpovídané o více kroků a zároveň na základě menšího množství informací.

4.3 Druhý experiment

Data jsme simulovali analogicky jako u prvního simulačního experimentu. Pouze jsme použili pro simulaci hodnot M_i binomické rozdělení s parametry $n = 120$ a $p = 3/4$. Pro modelování $K_m^{(i)}$ jsme tentokrát použili degenerované rozdělení $K_m^{(i)} = 8$ pro každé m .

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,91625 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,22069 & 1,45686 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,04007 & 1,19534 & 1,60001 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,09779 & 1,73453 & 2,03882 & 2,48813 \\ 0 & 0 & 0 & 0 & 0 & 1,40056 & 1,98486 & 2,19228 & 2,53734 & 3,02161 \\ 0 & 0 & 0 & 0 & 2,13510 & 2,92024 & 3,56938 & 3,84672 & 3,89370 & 4,12597 \\ 0 & 0 & 0 & 1,70107 & 2,62671 & 3,03796 & 3,25464 & 3,11827 & 3,37093 & 3,52980 \\ 0 & 0 & 3,91007 & 4,01789 & 4,36582 & 4,79687 & 4,91424 & 4,91698 & 5,08276 & 5,22427 \\ 0 & 5,05662 & 5,67632 & 5,81464 & 6,41503 & 6,58307 & 6,88285 & 6,76826 & 6,81932 & 7,00405 \end{pmatrix}$$

Tabulka 4.5: Průměrná relativní absolutní chyba predikce úhrnů počtů plateb v procentech spočítaná pomocí Mackova modelu

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,88629 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,12560 & 1,36530 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00149 & 1,16156 & 1,50741 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,07398 & 1,61462 & 1,85256 & 2,17708 \\ 0 & 0 & 0 & 0 & 0 & 1,38260 & 1,99496 & 2,17543 & 2,41572 & 2,72967 \\ 0 & 0 & 0 & 0 & 1,89939 & 2,54714 & 3,00628 & 3,11803 & 3,10769 & 3,26354 \\ 0 & 0 & 0 & 1,63077 & 2,47679 & 2,74553 & 2,80221 & 2,64416 & 2,72285 & 2,98348 \\ 0 & 0 & 3,61207 & 3,27620 & 3,49948 & 3,64935 & 3,65703 & 3,67796 & 3,77606 & 3,93441 \\ 0 & 4,68245 & 4,26377 & 4,57868 & 4,96066 & 4,99799 & 4,92184 & 4,82633 & 4,71054 & 4,74842 \end{pmatrix}$$

Tabulka 4.6: Průměrná relativní absolutní chyba predikce úhrnů počtů plateb v procentech spočítaná pomocí Poissonova shlukového modelu

Opětovně jsme takto obdrželi celý čtverec, na jehož horní levý trojúhelník

pohlížíme jako na „reálná“ data. Trojúhelník doplníme na čtverec pomocí predikcí získaných jak na základě Mackova modelu, tak Poissonova shlukového modelu.

Nejdříve se podíváme na průměrnou relativní absolutní chybu u úhrnu počtu plateb danou tabulkami 4.5 a 4.6. Opětovně i při tomto experimentu můžeme pozorovat, že Poissonův shlukový model dává menší průměrnou relativní absolutní chybu než Mackův model. Tentokrát je vidět, že už při předpovědi o jeden krok je tato chyba menší. Samozřejmě postupujeme-li směrem k pravému dolnímu rohu, chyba se v obou modelech zvětšuje. Průměrná relativní absolutní chyba je u Mackova modelu podstatně vyšší. V případě predikce o jeden krok roste chyba s klesajícím počtem vstupních informací.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,10630 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,27036 & 1,51079 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,35349 & 1,44938 & 1,76266 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,55339 & 2,13224 & 2,35308 & 2,80496 \\ 0 & 0 & 0 & 0 & 0 & 1,67532 & 2,14104 & 2,32891 & 2,75001 & 3,00664 \\ 0 & 0 & 0 & 0 & 2,14942 & 3,04975 & 3,70758 & 3,97359 & 3,94076 & 4,06251 \\ 0 & 0 & 0 & 2,37007 & 2,93590 & 3,56790 & 3,94674 & 3,96770 & 4,31534 & 4,42701 \\ 0 & 0 & 3,95162 & 4,12220 & 4,50469 & 4,92281 & 4,83503 & 4,73338 & 5,16470 & 5,07078 \\ 0 & 6,20406 & 7,30587 & 7,92135 & 8,46095 & 8,98091 & 9,06056 & 8,97690 & 8,94996 & 8,92136 \end{pmatrix}$$

Tabulka 4.7: Průměrná relativní absolutní chyba predikce úhrnů plateb za škody v procentech spočítaná pomocí Mackova modelu

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,98175 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,23410 & 1,37368 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,21361 & 1,32497 & 1,55145 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,37917 & 1,82016 & 2,14869 & 2,55828 \\ 0 & 0 & 0 & 0 & 0 & 1,55743 & 2,14319 & 2,29714 & 2,55136 & 2,82248 \\ 0 & 0 & 0 & 0 & 2,06155 & 2,81071 & 3,45535 & 3,53638 & 3,46483 & 3,59515 \\ 0 & 0 & 0 & 2,17586 & 2,70155 & 3,23696 & 3,29100 & 3,20827 & 3,17014 & 3,27282 \\ 0 & 0 & 3,96588 & 3,75183 & 3,98899 & 4,22470 & 4,31644 & 4,36203 & 4,42188 & 4,69351 \\ 0 & 5,54559 & 5,35009 & 5,88610 & 5,91254 & 6,13135 & 5,76413 & 5,57290 & 5,39667 & 5,38158 \end{pmatrix}$$

Tabulka 4.8: Průměrná relativní absolutní chyba predikce úhrnů plateb za škody v procentech spočítaná pomocí Poissonova shlukového modelu

Při porovnání tabulek 4.7 a 4.8 znázorňujících průměrnou relativní absolutní chybu úhrnů plateb za škody si všimneme, že pro tuto simulaci jsou v každém bodě průměrné absolutní relativní chyby Poissonova shlukového modelu menší než v Mackově modelu. Opětovně platí, že s postupem k pravému dolnímu rohu roste velikost chyby, což souvisí s výpočtem predikce o více kroků a na základě menšího počtu dat. Růst této chyby je rychlejší v případě Mackova modelu. V případě predikce o jeden krok opětovně vidíme, že chyba roste s klesajícím počtem počátečních informací.

4.4 Shrnutí

Naším cílem bylo porovnat dva způsoby predikce výše pojistných plnění, která budou vyplacena v budoucích letech za již vzniklé škody. První způsob vychá-

zí z Mackova modelu chain ladderu, druhý je založen na Poissonově shlukovém modelu. S rostoucím počtem kroků predikce a snižujícím se množstvím informací průměrná relativní absolutní chyba u obou metod roste avšak rozdílným tempem. Poissonův shlukový model vytváří znatelně menší chybu především pro predikci o více kroků dopředu a v situacích, kdy máme menší počet vstupních informací. V ostatních případech dávají obě metody srovnatelné výsledky. Poněkud překvapivé je, že metoda založená na Poissonově shlukovém modelu si vedla ve srovnání s Mackovým postupem lépe ve druhém experimentu. V něm byla data simulována z upraveného modelu, který přesně neodpovídal předpokladům použité metody. To ukazuje na jistou robustnost metody vzhledem k malým změnám předpokladů modelu. Z pohledu chybovosti je výhodnější postup vycházející z Poissonova shlukového modelu. Dalším důležitým aspektem je ovšem také jednoduchost. Mackův model potřebuje k výpočtu pouze informace o úhrnech počtů plateb a úhrnech plateb za jednotlivé roky do daného data, zatímco Poissonův shlukový model potřebuje ke svému výpočtu více informací v podobě počtu pojistných událostí a je početně podstatně náročnější.

Závěr

V této práci jsme se věnovali hlavně metodám, na jejichž základě jsme schopni odhadnout množství finančních prostředků, které si musí držet neživotní pojišťovna na škody již nastalé, ale dosud nenahlášené nebo na škody nastalé, nahlášené, ale dosud nezlikvidované.

V úvodu práce jsme se ve stručnosti seznámili s bodovými procesy, Poissonovým bodovým procesem, shlukovými bodovými procesy, kótovanými bodovými procesy a uvedli jsme si vazbu mezi bodovými procesy a pojistnou matematikou. Obecný rámec shlukových a kótovaných bodových procesů zahrnuje dva konkrétní modely, kterým jsme se podrobně věnovali v dalších dvou kapitolách. I když to nebylo nutné, použili jsme pro jejich definici jednotnou teorii bodových procesů.

Ve druhé kapitole jsme si ukázali metodu chain ladder a Mackův model. Uvedli jsme základní definice, chvíli jsme se věnovali asymptotickým vlastnostem metody chain ladder. Větší část kapitoly jsme věnovali odhadům vývojových faktorů, rozptylu, jejich momentům a vlastnostem. Klíčovou částí v této kapitole je tvorba jednokrokové a vícekové predikce a na jejím základě pak odvozená rezerva pro jednotlivé roky. Zajímala nás střední kvadratická chyba jednokrokové a vícekové predikce, která nám vypovídá o určité přesnosti tohoto modelu. Na závěr kapitoly jsme zařadili příklad na reálných datech, kde jsme si vyzkoušeli výpočet všech potřebných odhadů, včetně predikce úhrnů plateb, výpočtu rezervy a odhadu střední kvadratické chyby rezervy.

Mackův model předpokládá, že nejlepší predikce úhrnů počtů nebo výší plateb je násobek poslední hodnoty. Tento předpoklad podstatně zjednodušuje výpočty, ale není motivován žádným konkrétním modelem. Proto byl v [2] navržen jednoduchý stochastický model pro počty a příslušné výšky plateb. Tomuto modelu, který označujeme jako Poissonův shlukový model, jsme se věnovali ve třetí kapitole. Nejdříve jsme tento model definovali a s tím i všechny jeho proměnné. Posléze nás hlavně zajímala jednokroková a víceková predikce, kterou jsme odvodili pro obecnou situaci. Speciálně pokud má veličina M , která znázorňuje počet pojistných událostí v jednotlivých letech, Poissonovo, binomické nebo negativně binomické rozdělení, platí rekurzivní vzorce, které je možné použít k výpočtům predikce a není třeba využívat obecné vzorce, které v sobě mají zakomponovanou nekonečnou sumu. Na závěr této kapitoly jsme se krátce věnovali asymptotickému chování predikce.

U třetí kapitoly není zařazen výpočet na stejných reálných datech jako ve druhé kapitole, protože k tomuto výpočtu bychom potřebovali více informací než je pouhý kumulativní trojúhelník výší plateb. Proto je poslední kapitola věnována simulacím, na kterých jsme měli možnost napočítat obě metody. Zkusili jsme porovnat výsledky predikce úhrnů počtů plateb a úhrnů výší plateb. Provedli jsme dva různé simulační experimenty, které se lišily ve volbě rozdělení náhodných veličin M a K_m , kde první znamená počet pojistných událostí v daném období a druhá znázorňuje počet plateb k jednotlivým pojistným událostem. V prvním případě jsme volili obě rozdělení jako Poissonova a spočítali jsme průměrnou relativní absolutní chybu pro obě metody na základě padesáti simulací. Ukázalo se, že u obou metod roste chyba s klesajícím počtem výchozích dat a rostoucím krokem predikce, avšak u Poissonova shlukového modelu je tato chyba typicky

menší, i když při predikci o jeden či dva kroky byla v některých situacích mírně větší než u Mackova modelu. Při druhém simulačním experimentu jsme volili binomické rozdělení u veličiny M a degenerované rozdělení pro K_m . Metodu z třetí kapitoly jsme použili tak, že předpokládala Poissonovo rozdělení náhodných veličin M a K_m , což tentokrát na rozdíl od prvního experimentu nebylo splněno. Opět jsme napočítali průměrnou relativní absolutní chybu obou metod na základě padesáti opakování. Opětovně rostla u obou metod chyba při menším počtu známých dat a vyšším kroku predikce. V tomto případě byla i pro jednokrokovou predikci menší chyba u Poissonova shlukového modelu než u Mackova modelu, přičemž rozdíl byl znatelnější se zvětšujícím se krokem predikce a zvětšujícím se rokem i . U obou simulací bylo vidět, že chyba predikce o jeden krok roste se snižujícím se množstvím vstupních dat.

Z tohoto srovnání za pomoci průměrné relativní absolutní chyby by se mohlo zdát, že Poissonův shlukový model je výhodnější, protože dosahuje menší chybovosti. Je ale potřeba mít na paměti, že reálná data nebudou tak přesně odpovídat modelu jako naše simulovaná data. Dalším důležitým faktem je, že k jeho výpočtu potřebujeme více informací a že výpočet je mnohem náročnější než u Mackova modelu. Při volbě mezi těmito modely je třeba mít na paměti klady i zápory a rozhodnout se podle potřeby, co je v dané situaci důležitější, jestli menší chybovost nebo menší náročnost výpočtu.

Seznam použité literatury

- [1] Gould, I. L. (2008): *Stochastic chain-ladder models in nonlife insurance*, diplomová práce, University of Bergen.
- [2] Jessen, A. H., Mikosch, T., Samorodnitsky, G. (2011): Prediction of outstanding payments in a Poisson cluster model, *Scandinavian Actuarial Journal* **3**, 214–237.
- [3] Mack, T. (1993): Distribution-free calculation of the standard error of chain ladder reserve estimates, *ASTIN Bulletin* **23**, 213–225.
- [4] Mack, T. (1994): Measuring the variability of chain ladder reserve estimates, *In Proceedings of the Casualty Actuarial Society Spring Forum*, 101–182.
- [5] Mandl, P., Mazurová, L. (1999): *Matematické základy neživotního pojištění*, Matfyzpress, Praha.
- [6] Mikosch, T. (2009): *Non-Life Insurance Mathematics: An Introduction with the Poisson Process*, 2nd edition, Springer, Berlin.
- [7] Rataj, J. (2006): *Bodové procesy*, Karolinum, Praha.
- [8] Wolfram Research, Inc. (2010): *Wolfram Mathematica, verze 8.0*, Champaign.

Přílohy

V/Z	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0,700	6,000	17,100	21,557	27,980	30,980	34,236	38,795	43,551	47,320	48,807	50,881	51,496	52,805	52,646	52,695	52,695	52,699
2	2,000	7,700	13,654	26,334	38,734	48,492	52,685	57,677	61,664	62,646	64,204	66,754	68,493	69,570	70,418	70,217	70,462	
3	2,500	9,458	25,658	39,001	50,045	63,106	71,721	79,218	85,239	89,252	90,710	92,559	93,815	96,383	97,981	98,052		
4	2,688	11,958	25,241	38,562	53,836	70,607	83,326	92,246	100,784	104,161	107,791	111,542	114,025	116,686	117,068			
5	2,560	9,448	21,037	38,454	49,875	67,396	76,950	81,942	87,534	90,925	96,038	96,957	98,121	100,063				
6	2,795	10,220	27,252	54,768	81,070	100,423	115,359	118,617	121,727	126,110	132,068	131,688	131,801					
7	3,778	11,072	28,103	51,360	73,364	95,403	108,446	117,982	121,342	124,977	132,322	136,538						
8	5,083	15,237	35,488	65,717	92,923	106,712	116,284	132,801	146,637	149,505	151,035							
9	3,996	12,764	28,641	53,651	75,726	90,144	109,356	115,232	120,466	122,910								
10	5,451	17,508	39,835	67,750	89,010	112,132	123,015	134,509	142,955									
11	5,146	17,896	33,035	49,087	68,793	84,154	107,485	114,596										
12	8,264	19,681	29,160	44,865	63,715	80,831	93,745											
13	8,569	23,860	45,769	69,704	101,389	122,228												
14	7,088	21,831	33,254	55,304	75,132													
15	9,580	26,534	45,800	70,765														
16	9,172	32,395	48,358															
17	6,393	22,201																
18	6,423																	

Tabulka 4.9: Úhrny plateb za 18 let. Řádky znázorňují rok vzniku (V) pojistné události a sloupce zpoždění (Z) ve vyplacení plateb pojistného plnění.

V/Z	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0,700	6,000	17,100	21,557	27,980	30,980	34,236	38,795	43,551	47,320	48,807	50,881	51,496	52,805	52,646	52,695	52,695	52,699
2	2,000	7,700	13,654	26,334	38,734	48,492	52,685	57,677	61,664	62,646	64,204	66,754	68,493	69,570	70,418	70,217	70,462	70,467
3	2,500	9,458	25,658	39,001	50,045	63,106	71,721	79,218	85,239	89,252	90,710	92,559	93,815	96,383	97,981	98,052	98,247	98,255
4	2,688	11,958	25,241	38,562	53,836	70,607	83,326	92,246	100,784	104,161	107,791	111,542	114,025	116,686	117,068	117,025	117,258	117,267
5	2,560	9,448	21,037	38,454	49,875	67,396	76,950	81,942	87,534	90,925	96,038	96,957	98,121	100,063	100,859	100,822	101,023	101,031
6	2,795	10,220	27,252	54,768	81,070	100,423	115,359	118,617	121,727	126,110	132,068	131,688	131,801	134,758	135,830	135,781	136,051	136,062
7	3,778	11,072	28,103	51,360	73,364	95,403	108,446	117,982	121,342	124,977	132,322	136,538	138,366	141,471	142,596	142,544	142,828	142,839
8	5,083	15,237	35,488	65,717	92,923	106,712	116,284	132,801	146,637	149,505	151,035	154,402	156,469	159,980	161,253	161,194	161,515	161,528
9	3,996	12,764	28,641	53,651	75,726	90,144	109,356	115,232	120,466	122,910	127,252	130,088	131,830	134,788	135,861	135,811	136,082	136,092
10	5,451	17,508	39,835	67,750	89,010	112,132	123,015	134,509	142,955	147,596	152,810	156,217	158,308	161,860	163,148	163,088	163,414	163,426
11	5,146	17,896	33,035	49,087	68,793	84,154	107,485	114,596	122,032	125,994	130,445	133,353	135,139	138,171	139,270	139,219	139,496	139,507
12	8,264	19,681	29,160	44,865	63,715	80,831	93,745	101,699	108,298	111,815	115,764	118,345	119,930	122,621	123,596	123,551	123,797	123,807
13	8,569	23,860	45,769	69,704	101,389	122,228	140,520	152,443	162,335	167,605	173,526	177,394	179,770	183,803	185,266	185,198	185,567	185,581
14	7,088	21,831	33,254	55,304	75,132	93,007	106,926	115,999	123,526	127,537	132,042	134,985	136,793	139,862	140,975	140,923	141,204	141,215
15	9,580	26,534	45,800	70,765	98,551	121,998	140,256	152,156	162,030	167,291	173,200	177,061	179,432	183,458	184,918	184,850	185,218	185,232
16	9,172	32,395	48,358	80,435	112,018	138,670	159,422	172,949	184,172	190,151	196,868	201,257	203,952	208,528	210,187	210,110	210,529	210,545
17	6,393	22,201	43,549	72,437	100,879	124,880	143,569	155,751	165,858	171,243	177,292	181,244	183,671	187,792	189,286	189,217	189,594	189,608
18	6,423	20,653	40,512	67,385	93,843	116,170	133,556	144,888	154,290	159,299	164,926	168,603	170,860	174,694	176,084	176,019	176,370	176,384

Tabulka 4.10: Úhrny plateb za 18 let. Řádky znázorňují rok vzniku (V) pojistné události a sloupce zpoždění (Z) ve vyplacení plateb pojistného plnění. Původní trojúhelník je doplněn na čtverec pomocí příslušných predikcí.