Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Veronika Dlouhá

Výpočty distribuce agregovaných škod

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce:	RNDr. Michal Pešta, Ph.D.
Studijní program:	Matematika
Studijní obor:	Obecná matematika

Praha 2015

Prohlašuji, že jsem tuto bakalářskou 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 .... dne .... Podpis autora

Poděkování

Ráda bych poděkovala svému vedoucímu RNDr. Michalovi Peštovi, Ph.D. za jeho odborné vedení, cenné připomínky a rady při zpracování této práce.

Název práce: Výpočty distribuce agregovaných škod

Autor: Veronika Dlouhá

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Michal Pešta, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V této bakalářské práci se budu zabývat rozdělením celkových škod. Nejdříve představím danou problematiku a uvedu základní modely. Následně ukáži některá rozdělení používaná pro modelování počtu a výše škod i s odhady jejich parametrů. Poté se dostanu ke složenému rozdělení a uvedu jeho základní vlastnosti. V dalších částech práce budu zkoumat pravděpodobnost výše celkové škody pomocí Panjerovy rekurze a rychlé Fourierovy transformace a aplikuji získanou teorii na příkladech. Na závěr uvedu některé metody aproximace celkové škody pomocí známých rozdělení.

Klíčová slova: agregované škody, kolektivní model, Panjerova rekurze, rychlá Fourierova transformace

Title: Computing the aggregate claims distribution

Author: Veronika Dlouhá

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Michal Pešta, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In this bachelor thesis I study the distribution of aggregate claims. First I introduce the topic and present main models. Then, I show some distributions used for modeling number and amound of claims with estimates of their parameters. Next I get to compound distribution and its basic charakteristics. In other parts of the thesis I study the probability of aggregate claims using Panjer recursion and fast Fourier transform and apply the thoery to examples. Finally I mention some methods to approximate aggregate claims using known distribution.

Keywords: aggregate claims, collective model, Panjer recursion, fast Fourier transform

Obsah

1 Úvod **2**

2 Rozdělení počtu škod **3**
2.1 Poissonovo rozdělení 3
2.2 Poissonovo rozdělení s náhodným parametrem 4

3 Rozdělení výše škod **5**
3.1 Exponenciální rozdělení 5
3.2 Paretovo rozdělení 5
3.3 Gamma a Logaritmicko-normální rozdělení 5

4 Odhady parametrů rozdělení výše a počtu škod **6**
4.1 Poissonovo rozdělení 6
4.2 Negativně binomické rozdělení 7
4.3 Exponenciální rozdělení 8
4.4 Gamma rozdělení 8
4.5 Logaritmicko-normální rozdělení 9
4.6 Paretovo rozdělení 9

5 Složené rozdělení **11**
5.1 Střední hodnota a rozptyl 11
5.2 Momentová vytvořující funkce 12
5.3 Charakteristická funkce 13
5.4 Součet složených Poissonových rozdělení 13

6 Panjerova rekurze **15**
6.1 Rozdělení vhodná pro použití Panjerovy rekurze 16
6.2 Aplikace Panjerovy rekurze 17

7 Rychlá Fourierova transformace **20**
7.1 Aplikace Fourierovy transformace 21

8 Aproximace celkové škody **23**
8.1 Aproximace normálním rozdělením 23
8.2 Aproximace logaritmicko-normálním rozdělením 24
8.3 Příklad 24

9 Závěr **26**

Seznam použité literatury **27**

Kapitola 1

Úvod

Cílem této bakalářské práce je určit rozdělení agregovaných škod. K tomu potřebujeme vybudovat model celkové výše škod, které nastaly během určitého období. Základním kamenem jsou náhodné veličiny popisující počet škod v určitém období a výši jednotlivých škod.

Máme dva základní druhy modelování rizika: kolektivní a individuální model. První z nich počítá s jednotlivými škodami samostatně, nehledě na pojistné smlouvy. V tomto případě můžeme celkovou výši škod určit jako součet N individuálních škod X_i, kde N a X_i jsou náhodné veličiny. Tedy

$$S = X_1 + X_2 + \cdots + X_N, N = 0,\ 1,\ 2, \ldots, \tag{1.1}$$

kde $S = 0$ pokud $N = 0$.

Definice 1. Kolektivní model rizika *má reprezentaci jako v* (1.1)*, kde X_i jsou stejně rozdělené náhodné veličiny a náhodná veličina N nezávisí na X_i, $i = 1{,}2, \ldots$.*

Druhý model počítá s jednotlivými pojistnými smlouvami.

Definice 2. Individuální model rizika *reprezentuje celkovou škodu jako součet $S = X_1 + \cdots + X_n$, kde n je pevně stanovený počet pojistných smluv a X_i je náhodná veličina reprezentující škodu pro i-tou pojistnou smlouvu. Předpokládáme, že $X_1, \ldots, X_n$ jsou nezávislé.*

Na rozdíl od kolektivního modelu u individuálního modelu nepředpokládáme stejné rozdělení výše škod X_i. V případě, že výše škod jsou stejně rozdělena, je individuální model speciálním případem kolektivního, kde N má degenerované rozdělení $\mathsf{P}(N = n) = 1$.

V dalších kapitolách se budeme zabývat pouze kolektivním modelem.

Kapitola 2

Rozdělení počtu škod

V této kapitole představíme nejčastěji užívaná rozdělení k modelování celkového počtu škod N v kolektivním modelu. Tato rozdělení jsou diskrétní a nenabývají záporných hodnot.

2.1 Poissonovo rozdělení

První volbou pro rozdělení počtu škod je Poissonovo rozdělení $Po(\lambda)$, které je definované pravděpodobnostmi

$$\mathsf{P}(N = n) = \frac{\lambda^n}{n!}e^{-\lambda}, n = 0,1,\ldots,\lambda > 0.$$

K tomuto rozdělení se dostaneme tak, že nejprve předpokládáme individuální model z definice 2 o ν pojistných smlouvách, kde $X_1,\ldots,X_\nu$ jsou nezávislé a stejně rozdělené a platí, že

$$\mathsf{P}(X_i \neq 0) = p, i = 1,\ldots,\nu,\ p \in (0,1]$$

(pravděpodobnost že v rámci jedné pojistné smlouvy dojde ke škodě).

Pokud budeme předpokládat, že v rámci jedné pojistné smlouvy nemůže nastat více pojistných událostí, jsou náhodné veličiny $X_1,\ldots,X_N$ kolektivního modelu představovány nenulovými hodnotami z individuálního modelu. Potom pro rozdělení počtu škod N platí

$$\mathsf{P}(N = n) = \binom{\nu}{n} p^n (1-p)^{(\nu-n)}, n = 0,\ldots,\nu, p \in (0,1],$$

tedy N má binomické rozdělení s parametrem p.

Prakticky je pravděpodobnost p malá a počet pojistných smluv ν velký, tedy přejdeme k limitě $p \to 0, \nu \to \infty, \nu p \to \lambda$. Potom

$$\mathsf{P}(N = n) = \frac{\nu(\nu-1)\ldots(\nu-n+1)}{n!} p^n \left(1 - \frac{\lambda}{\nu}\right)^{(\nu-n)} \to \frac{\lambda^n}{n!}e^{-\lambda}, n = 0,1,\ldots.$$

Tedy N má limitně Poissonovo rozdělení.

Pro Poissonovo rozdělení platí $\mathsf{E}[X] = \mathrm{var}[X] = \lambda$, tedy Poissonovo rozdělení nemůžeme použít, pokud rozptyl výrazně převyšuje střední hodnotu.

2.2 Poissonovo rozdělení s náhodným parametrem

Podle Kaas a kol. (2008) se můžeme pomocí Poissonova rozdělení s náhodným parametrem dostat k negativně binomickému rozdělení následujícím způsobem.

Předpokládejme, že každá pojistná smlouva má počet škod modelován Poissonovým rozdělením, přitom parametr $\lambda > 0$ je neznámý a různý pro každou pojistnou smlouvu. Předpokládejme tedy, že parametr λ je výsledkem náhodné veličiny Λ s distribuční funkcí U. Rozdělení počtu škod za podmínky, že $\Lambda = \lambda$, je Poissonovo s parametrem λ a platí

$$\mathsf{P}(N = n) = \int_0^\infty \mathsf{P}(N = n|\Lambda = \lambda)dU(\lambda) = \int_0^\infty \frac{\lambda^n}{n!}e^{-\lambda}dU(\lambda).$$

Navíc platí (z vlastností podmíněné střední hodnoty a podmíněného rozptylu):

$$\begin{aligned} \mathsf{E}[N] &= \mathsf{E}[\mathsf{E}[N|\Lambda]] = \mathsf{E}[\Lambda], \\ \mathrm{var}[N] &= \mathsf{E}[\mathrm{var}[N|\Lambda]] + \mathrm{var}[\mathsf{E}[N|\Lambda]] = \mathsf{E}[\Lambda] + \mathrm{var}[\Lambda] \geq \mathsf{E}[N], \end{aligned}$$

tedy, na rozdíl od předchozího modelu, tento model můžeme použít, i když rozptyl převyšuje střední hodnotu.

Nyní předpokládejme navíc, že Λ má gamma rozdělení $\Gamma(\alpha,\beta)$, $\alpha > 0$, $\beta > 0$. V následujícím budeme využívat vlastnosti podmíněné střední hodnoty ($\mathsf{E}[e^{tN}] = \mathsf{E}[\mathsf{E}[e^{tN}|\Lambda]]$). Potom momentová vytvořující funkce celkového počtu škod má tvar:

$$\begin{aligned} m_N(t) &= \mathsf{E}[e^{tN}] = \mathsf{E}[\mathsf{E}[e^{tN}|\Lambda]] = \mathsf{E}[e^{\Lambda(e^t-1)}] = m_\Lambda(e^t - 1) \\ &= \left(\frac{\beta}{\beta - (e^t - 1)}\right)^\alpha = \left(\frac{\frac{\beta}{\beta+1}}{1 - e^t(1 - \frac{\beta}{\beta+1})}\right)^\alpha, \end{aligned}$$

ve které můžeme poznat momentovou vytvořující funkci negativně binomického rozdělení $NB(\alpha, \frac{\beta}{\beta+1})$.

Kapitola 3

Rozdělení výše škod

V této kapitole představíme nejčastěji užívaná rozdělení k modelování výše jednotlivých škod v kolektivním modelu. Tato rozdělení jsou spojitá a nenabývají záporných hodnot.

3.1 Exponenciální rozdělení

První volbou pro rozdělení počtu škod je Exponenciální rozdělení ($Exp(\beta)$), které je definované hustotou:

$$f(x) = \frac{1}{\beta} e^{-\frac{1}{\beta}x}, \qquad 0 \leq x < \infty, \quad 0 < \beta.$$

3.2 Paretovo rozdělení

Dalším z často používaných rozdělení je Paretovo rozdělení ($Pareto(\alpha,\beta)$), které je definované hustotou:

$$f(x) = \frac{\beta \alpha^{\beta}}{x^{\beta+1}} \qquad \alpha \leq x < \infty, \quad 0 < \alpha, \quad 0 < \beta.$$

Toto rozdělení se používá, pokud je vysoká pravděpodobnost vyšších škod.

3.3 Gamma a Logaritmicko-normální rozdělení

Poslední rozdělení, které zmíním jsou Gamma a Logaritmicko-normální rozdělení. Gamma ($\Gamma(\alpha,\beta)$) rozdělení je definované hustotou:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \qquad 0 \leq x < \infty, \quad 0 < \alpha, \quad 0 < \beta.$$

Logaritmicko-normální ($LN(\mu,\sigma^2)$) rozdělení je definované hustotou:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \qquad 0 < x.$$

Kapitola 4

Odhady parametrů rozdělení výše a počtu škod

Jestliže má náhodná veličina Y rozdělení, které závisí na parametru $\theta = (\theta_1,\theta_2,\ldots,\theta_n)$, snažíme se získat odhad tohoto parametru tak, aby měl co nejlepší vlastnosti.

Hlavní metodou odhadu je metoda maximální věrohodnosti, která maximalizuje věrohodnostní funkci: $\mathcal{L}(\vec{y},\theta) = \prod_{i=1}^{n} f_{Y_i}(y_i,\theta)$, kde $Y_1,\ldots,Y_n$ je náhodný výběr z rozdělení, z kterého pochází náhodná veličina Y. Ekvivalentní řešení dostaneme i maximalizováním logaritmu věrohodnostní funkce, což bývá většinou jednodušší. Označíme $\ell(\vec{y},\theta) = log(\mathcal{L}(\vec{y},\theta))$.

Jestliže je funkce ℓ spojitá v θ maximalizujeme ji pomocí soustavy rovnic:

$$
\begin{aligned}
\frac{\partial \ell}{\partial \theta_1} &\stackrel{!}{=} 0 \\
\frac{\partial \ell}{\partial \theta_2} &\stackrel{!}{=} 0 \\
&\vdots \\
\frac{\partial \ell}{\partial \theta_n} &\stackrel{!}{=} 0.
\end{aligned}
$$

Často při výpočtu odhadů používáme iterační metody, pro které potřebujeme inicializační hodnoty. Ty můžeme získat metodou momentů. Momentová metoda spočívá v tom, že porovnáme výběrové momenty získaných dat s teoretickými momenty předpokládaného rozdělení. Tím opět vzniká soustava rovnic.

4.1 Poissonovo rozdělení

Nechť $Y_1,\ldots,Y_n$ je náhodný výběr z Poissonova rozdělení s parametrem λ. Potom platí:

$$
\begin{aligned}
\ell(\vec{y},\lambda) &= \log \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} \\
&= -\lambda n + \sum_{i=1}^{n} y_i log\lambda - \sum_{i=1}^{n} \log y_i!.
\end{aligned}
$$

Pomocí parciální derivace zjistíme maximální věrohodný odhad $\hat{\lambda}$:

$$\frac{\partial \ell}{\partial \lambda} = -n + \sum \frac{y_i}{\hat{\lambda}} \overset{!}{=} 0$$
$$\hat{\lambda} = \frac{1}{n} \sum y_i$$
$$\hat{\lambda} = \bar{y}.$$

4.2 Negativně binomické rozdělení

Nechť $Y_1, \ldots, Y_n$ je náhodný výběr z Negativně binomického rozdělení s parametry r a p. Potom platí:

$$\ell(\vec{y}, p, r) = \log \prod_{i=1}^{n} \binom{r + y_i - 1}{y_i} p^r (1-p)^{y_i}$$
$$= \sum_{i=1}^{n} \sum_{m=0}^{y_i - 1} \log(r+m) - \log(y_i!) + nr \log p + \sum_{i=1}^{n} y_i \log(1-p).$$

Pomocí parciálních derivací zjistíme maximální věrohodné odhady $\hat{p}$ a $\hat{r}$:

$$\frac{\partial \ell}{\partial p} = \frac{nr}{p} - \frac{1}{1-p} \sum_{i=1}^{n} y_i \overset{!}{=} 0$$
$$\frac{1}{1-\hat{p}} \bar{y} = \frac{\hat{r}}{\hat{p}}$$
$$\hat{p} = \frac{\hat{r}}{\hat{r} + \bar{y}}.$$

Odhad $\hat{p}$ můžeme dosadit do parciální derivace podle r. Tím získáme jednu rovnici o jedné neznámé, která se dá řešit iteračně pomocí některého z matematických softwarů.

$$\frac{\partial \ell}{\partial r} = n \log p + \sum_{i=1}^{n} \sum_{m=0}^{y_i - 1} \frac{1}{r+m} \overset{!}{=} 0$$
$$n \log \frac{\hat{r}}{\hat{r} + \bar{y}} + \sum_{i=1}^{n} \sum_{m=0}^{y_i - 1} \frac{1}{\hat{r} + m} = 0.$$

Pro získání inicializačních hodnot iterační metody použijeme metodu momentů:

$$\frac{r(1-p)}{p} \overset{!}{=} \bar{y},$$
$$\frac{r(1-p)}{p^2} \overset{!}{=} \bar{y^2} - \bar{y}^2.$$

Vydělením první rovnice druhou, získáme řešení p_0, a dosazením do první rovnice řešení r_0:

$$p_0 = \frac{\bar{y}}{\bar{y^2} - \bar{y}^2},$$
$$r_0 = \frac{\bar{y}^2}{\bar{y^2} - \bar{y}^2 - \bar{y}}.$$

.

4.3 Exponenciální rozdělení

Nechť $Y_1, \ldots, Y_n$ je náhodný výběr z Exponenciálního rozdělení s parametrem β. Potom platí:

$$\begin{aligned}\ell(\vec{y},\alpha,\beta) &= \log \prod_{i=1}^{n} \frac{1}{\beta} e^{-\frac{1}{\beta} y_i} \\ &= -n \log \beta - \frac{1}{\beta} \sum_{i=1}^{n} y_i.\end{aligned}$$

Pomocí parciální derivace zjistíme maximální věrohodný odhad $\hat{\beta}$:

$$\begin{aligned}\frac{\partial \ell}{\partial \beta} = \frac{-n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} y_i &\stackrel{!}{=} 0 \\ -n + \frac{1}{\beta} \sum_{i=1}^{n} y_i &= 0 \\ \hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} y_i &= \bar{y}.\end{aligned}$$

4.4 Gamma rozdělení

Nechť $Y_1, \ldots, Y_n$ je náhodný výběr z Gamma rozdělení s parametry α a β. Potom platí:

$$\begin{aligned}\ell(\vec{y},\alpha,\beta) &= \log \prod_{i=1}^{n} \frac{y_i^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}} \\ &= (\alpha - 1) \sum_{i=1}^{n} \log y_i - \frac{1}{\beta} \sum_{i=1}^{n} y_i - n \log \Gamma(\alpha) - n\alpha \log \beta.\end{aligned}$$

Pomocí parciálních derivací zjistíme maximální věrohodné odhady $\hat{\alpha}$ a $\hat{\beta}$:

$$\begin{aligned}\frac{\partial \ell}{\partial \beta} &= \frac{1}{\beta^2} \sum_{i=1}^{n} y_i - \frac{n\alpha}{\beta} \stackrel{!}{=} 0 \\ \hat{\beta} &= \frac{\sum_{i=1}^{n} y_i}{n\hat{\alpha}} \\ \hat{\beta} &= \frac{1}{\alpha} \bar{y}.\end{aligned}$$

Odhad $\hat{\beta}$ můžeme dosadit do parciální derivace podle α. Tím získáme jednu rovnici o jedné neznámé, která se dá řešit numericky pomocí některého z matematických softwarů.

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n} \log y_i - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \log \beta &\stackrel{!}{=} 0 \\ \sum_{i=1}^{n} \log y_i - n \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} - n \log \left(\frac{1}{\hat{\alpha}} \bar{y} \right) &= 0.\end{aligned}$$

Pro získání inicializačních hodnot iterační metody použijeme metodu momentů:

$$\alpha\beta \stackrel{!}{=} \bar{y},$$
$$\alpha\beta^2 \stackrel{!}{=} \bar{y^2} - \bar{y}^2.$$

Vydělením druhé rovnice první získáme odhad pro parametr β. A dosazením do první rovnice, získáme odhad pro parametr α.

$$\hat{\beta} = \frac{\bar{y^2} - \bar{y}^2}{\bar{y}},$$
$$\hat{\alpha} = \frac{\bar{y}^2}{\bar{y^2} - \bar{y}^2}.$$

4.5 Logaritmicko-normální rozdělení

Nechť $Y_1, \ldots, Y_n$ je náhodný výběr z Logaritmicko-normálního rozdělení s parametry μ a σ^2. Potom platí:

$$\ell(\vec{y},\mu,\sigma^2) = \log\left(\prod_{i=1}^{n} \frac{e^{-(\log(y_i)\,-\,\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}y_i}\right)$$
$$= -\sum_{i=1}^{n} \log(\sqrt{2\pi}y_i) - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n} \frac{(\log(y_i) - \mu)^2}{2\sigma^2}.$$

Pomocí parciálních derivací zjistíme maximální věrohodné odhady parametrů $\hat{\mu}$ a $\hat{\sigma}^2$:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} 2\log y_i - n\mu\right) \stackrel{!}{=} 0,$$
$$\frac{\partial \ell}{\partial \sigma^2} = \frac{1}{2\sigma^2}\left(-n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(\log y_i - \mu)^2\right) \stackrel{!}{=} 0.$$

Z první rovnice ihned dostáváme odhad parametru μ a po dosazení do druhé rovnice i odhad parametru σ^2:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} 2\log y_i,$$
$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(\log y_i - \frac{1}{n}\sum_{k=1}^{n} 2\log y_k)^2.$$

4.6 Paretovo rozdělení

Nechť $Y_1, \ldots, Y_n$ je náhodný výběr z Paretova rozdělení s parametry α a β. Označme si $y_{(1)} = \min y_i$ Potom platí

$$\ell(\vec{y},\alpha,\beta) = \begin{cases} n\log\beta + n\beta\log\alpha - (\beta+1)\sum_{i=1}^{n}\log y_i & \alpha \le y_{(1)}, \\ -\infty & \alpha > y_{(1)}. \end{cases}$$

Pro libovolné $\beta > 0$ platí $\ell(\vec{y},\alpha,\beta) \leq \ell(\vec{y},y_{(1)},\beta)$, tedy maximálně věrohodný odhad pro parametr α musí být $\hat{\alpha} = y_{(1)}$.

Maximálně věrohodný odhad parametru β získáme pomocí parciální derivace logaritmické věrohodnostní funkce a použijeme již zjištěný odhad $\hat{\alpha}$:

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} + n \log \hat{\alpha} - \sum_{i=1}^{n} \log y_i \stackrel{!}{=} 0.$$

$$\begin{aligned} \hat{\beta} &= \left(\frac{1}{n} \sum_{i=1}^{n} \log y_i - \log \hat{\alpha} \right)^{-1} \\ &= \left(\frac{1}{n} \sum_{i=1}^{n} \log \frac{y_i}{y_{(1)}} \right)^{-1}. \end{aligned}$$

Kapitola 5

Složené rozdělení

Složené rozdělení, je rozdělení kolektivního modelu (1) $S = \sum_{i=1}^{N} X_i$. V této kapitole se budeme zabývat základními charakteristikami a vlastnostmi složeného rozdělení.

Pro jednodušší zápis si označme X jako náhodnou veličinu, která má stejné rozdělení jako X_i.

V celé kapitole budeme využívat podmíněnou střední hodnotu, pro kterou platí, že jestliže jsou X a Y náhodné veličiny s konečnou střední hodnotou, potom $\mathsf{E}\left[X\right] = \mathsf{E}\left[\mathsf{E}\left[X|Y\right]\right]$

5.1 Střední hodnota a rozptyl

Často se setkáváme s tím, že potřebujeme znát očekávanou hodnotu celkové škody, nebo chceme vědět, jak moc jsou hodnoty celkové škody rozptýleny. Tedy chceme znát střední hodnotu a rozptyl náhodné veličiny S.

Z vlastnosti podmíněné střední hodnoty platí:

$$\mathsf{E}\left[S\right] = \mathsf{E}\left[\mathsf{E}\left[S|N\right]\right].$$

N má diskrétní rozdělení, tedy

$$\mathsf{E}\left[\mathsf{E}\left[S|N\right]\right] = \sum_{n=0}^{\infty} \mathsf{E}\left[X_1 + \cdots + X_N | N = n\right] \mathsf{P}(N = n).$$

Z teorie pravděpodobnosti víme, že pokud jsou Y a Z, nezávislé náhodné veličiny a $G(Y,Z)$ je reálná měřitelná funkce potom $\mathsf{E}\left[G(Y,Z)|Z = z\right] = \mathsf{E}\left[G(Y,z)\right]$. Tedy

$$E[X_1 + \cdots + X_N | N = n] = E[X_1 + \cdots + X_n] = nE[X].$$

Potom platí:

$$\mathsf{E}\left[S\right] = \mathsf{E}\left[X\right] \sum_{n=0}^{\infty} n\, \mathsf{P}(N = n) = \mathsf{E}\left[X\right] \mathsf{E}\left[N\right] \tag{5.1}$$

Z vlastnosti podmíněného rozptylu platí:

$$\operatorname{var} S = \mathsf{E}\left[\operatorname{var}[S|N]\right] + \operatorname{var}[\mathsf{E}\left[S|N\right]].$$

Rovnost $\mathsf{E}[S|N] = N\,\mathsf{E}[X]$ platí z důkazu (5.1). Nyní podobně dokážeme, že $\mathrm{var}[S|N] = N\,\mathrm{var}[X]$:

$$\mathrm{var}[S|N] = \mathsf{E}[(\sum_{i=1}^{N} X_i - EX)^2|N].$$

Z nezávislosti X_i plyne

$$\mathsf{E}[(\sum_{i=1}^{N} X_i - EX)^2|N] = \mathsf{E}[\sum_{i=1}^{N}(X_i - EX)^2|N].$$

Nyní postupujeme podobně jako při důkazu (5.1):

$$\begin{aligned}\mathsf{E}[\sum_{i=1}^{N}(X_i - EX)^2|N] &= \sum_{n=0}^{\infty} E[\sum_{i=1}^{N}(X_i - EX)^2|N = n]\,\mathsf{P}(N = n)\\ &= \sum_{n=0}^{\infty} E[\sum_{i=1}^{n}(X_i - EX)^2]\,\mathsf{P}(N = n)\\ &= \sum_{n=0}^{\infty} n\,\mathrm{var}(X)\,\mathsf{P}(N = n) = N\,\mathrm{var}[X].\end{aligned}$$

Tedy platí rovnost:

$$\mathrm{var}\,S = \mathsf{E}[N\,\mathrm{var}[X]] + \mathrm{var}[N\,\mathsf{E}[X]] = \mathsf{E}[N]\,\mathrm{var}[X] + (\mathsf{E}[X])^2\,\mathrm{var}[N]. \tag{5.2}$$

5.2 Momentová vytvořující funkce

V dalších kapitolách se nám bude hodit momentová vytvořující funkce složeného rozdělení, jelikož jednoznačně určuje rozdělení.

Momentovou vytvořující funkci můžeme také využít k určení momentů složeného rozdělení.

Z definice a využití vlastnosti podmíněné střední hodnoty platí:

$$m_S(t) = \mathsf{E}[e^{tS}] = \mathsf{E}[\mathsf{E}[e^{tS}|N]].$$

Náhodná veličina N je diskrétní, tedy

$$\mathsf{E}[\mathsf{E}[e^{tS}|N]] = \sum_{n=0}^{\infty} \mathsf{E}[e^{t(X_1+\cdots+X_N)}|N = n]\,\mathsf{P}(N = n).$$

Z teorie pravděpodobnosti víme, že pokud jsou Y a Z, nezávislé náhodné veličiny a $G(Y,Z)$ je reálná měřitelná funkce potom $\mathsf{E}[G(Y,Z)|Z = z] = \mathsf{E}[G(Y,z)]$. Tedy

$$\mathsf{E}[e^{t(X_1+\cdots+X_N)}|N = n] = \mathsf{E}[e^{t(X_1+\cdots+X_n)}].$$

$X_1,\ldots,X_n$ jsou nezávislé a stejně rozdělené, potom i náhodné veličiny $e^{tX_1},\ldots,e^{tX_n}$ jsou nezávislé a stejně rozdělené a platí:

$$\mathsf{E}[e^{t(X_1+\cdots+X_n)}] = (\mathsf{E}[e^{t(X)}])^n = (m_X(t))^n.$$

Funkci $(m_X(t))^n$ můžeme přepsat, jako$(e^{\log m_X(t)})^n$. Tedy platí vzorec:

$$\begin{aligned}m_S(t) &= \sum_{n=0}^{\infty}(e^{\log m_X(t)})^n\,\mathsf{P}(N = n)\\ &= \mathsf{E}[(e^{\log m_X(t)})^N] = m_N(\log m_X(t)).\end{aligned} \tag{5.3}$$

5.3 Charakteristická funkce

V kapitole 7 budeme potřebovat charakteristickou funkci složeného rozdělení S.

Z definice charakteristické funkce a vlastnosti podmíněné střední hodnoty platí:

$$\phi_S(t) = \mathsf{E}\left[e^{itS}\right] = \mathsf{E}\left[\mathsf{E}\left[e^{itS}|N\right]\right]$$

N má diskrétní rozdělení a tedy platí

$$\mathsf{E}\left[\mathsf{E}\left[e^{itS}|N\right]\right] = \sum_{n=0}^{\infty} \mathsf{E}\left[e^{it(X_1+\cdots+X_N)}|N=n\right] \mathsf{P}(N=n)$$

Z teorie pravděpodobnosti víme, že pokud jsou Y a Z, nezávislé náhodné veličiny a $G(Y,Z)$ je reálná měřitelná funkce potom $\mathsf{E}\left[G(Y,Z)|Z=z\right] = \mathsf{E}\left[G(Y,z)\right]$. Tedy

$$\mathsf{E}\left[e^{it(X_1+\cdots+X_N)}|N=n\right] = \mathsf{E}\left[e^{it(X_1+\cdots+X_n)}\right].$$

$X_1,\ldots,X_n$ jsou nezávislé a stejně rozdělené, potom i náhodné veličiny $e^{itX_1},\ldots,e^{tX_n}$ jsou nezávislé a stejně rozdělené a platí:

$$\mathsf{E}\left[e^{it(X_1+\cdots+X_n)}\right] = \left(\mathsf{E}\left[e^{itX}\right]\right)^n = (\phi_X(t))^n.$$

Tedy platí rovnost:

$$\begin{aligned}\phi_S(t) &= \sum_{n=0}^{\infty} (\phi_X(t))^n \,\mathsf{P}(N=n)\\ &= \mathsf{E}\left[(\phi_X(t))^N\right] = P_N(\phi_X(t)),\end{aligned}$$

kde $P_N(t)$ je pravděpodobnostní vytvořující funkce náhodné veličiny N.

5.4 Součet složených Poissonových rozdělení

Řekneme, že $S = \sum_{i=1}^{N} X_i$ má složené Poissonovo rozdělení, jestliže náhodná veličina N má Poissonovo rozdělení.

Věta 1. *Nechť $S_1,\ldots,S_m$ jsou nezávislé náhodné veličiny se složeným Poissonovým rozdělením s parametrem λ_i a distribuční funkcí výše pojistných škod F_i, $i = 1,\ldots,m$. Potom $S = S_1 + \cdots + S_m$ je složené Poissonovo rozdělení s parametrem $\lambda = \sum_{i=1}^{m} \lambda_i$ a distribuční funkcí výše pojistných škod $F = \sum_{i=1}^{m} \frac{\lambda_i}{\lambda} F_i$*

Důkaz této věty je převzat z (Klugman a kol., 2012). Uvádím ho zde proto, že je zde vidět využití momentové vytvořující funkce k určení rozdělení.

Důkaz. Označíme m_{x_i} momentovou vytvořující funkci rozdělení s distribuční funkcí F_i. S využitím rovnosti (5.3) (použijeme na náhodné veličiny S_i) a znalostí momentové vytvořující funkce Poissonova rozdělení spočteme momentovou vytvořující funkci náhodné veličiny S.

$$\begin{aligned}m_S(t) &= \mathsf{E}\left[e^{t(S_1+\cdots+S_m)}\right] = \Pi_{i=1}^{m} m_{S_i} = \Pi_{i=1}^{m} e^{\lambda_i(m_{x_i}-1)}\\ &= e^{\lambda\left(\sum_{i=1}^{m} \frac{\lambda_i}{\lambda} m_{x_i} - \sum_{i=1}^{m} \frac{\lambda_i}{\lambda}\right)}\\ &= e^{\lambda\left(\sum_{i=1}^{m} \frac{\lambda_i}{\lambda} m_{x_i} - 1\right)}\end{aligned}$$

Jelikož $\sum_{i=1}^{m} \frac{\lambda_i}{\lambda} m_{x_i}$ je momentová vytvořující funkce rozdělení s distribuční funkcí $F = \sum_{i=1}^{m} \frac{\lambda_i}{\lambda} F_i$, má S složené Poissonovo rozdělení podle zadání.

❑

Podobné věty existují i pro binomické a negativně binomické rozdělení, které můžeme i s důkazem najít například v Koudelka (2014)

Kapitola 6

Panjerova rekurze

Panjerova rekurze je rekurzivní metoda výpočtu $f(x) = P(S = x)$, která však klade určité podmínky na rozdělení počtu a výše škod.

Věta 2. *(Panjerova formule)*
Nechť $S = \sum_{i=1}^{N} X_i$ je složené rozdělení, kde velikosti škod X_i jsou nezáporné celočíselné náhodné veličiny s pravděpodobnostní funkcí $p(x) = \mathsf{P}(X_i = x)$ a počet škod N je náhodná veličina, pro jejíž pravděpodobnostní funkci $q(n) = \mathsf{P}(N = n)$ platí

$$q(n) = \left(a + \frac{b}{n}\right) q(n-1), \quad n = 1{,}2, \ldots, \tag{6.1}$$

kde a,b jsou libovolná reálná čísla. Potom pro pravděpodobnostní funkci složeného rozdělení $f(x) = \mathsf{P}(S = x)$ platí následující vztahy:

$$f(0) = \begin{cases} \mathsf{P}(N = 0), & \text{je-li } p(0) = 0, \\ m_N(\log p(0)), & \text{je-li } p(0) > 0 \end{cases} \tag{6.2}$$

$$f(s) = \frac{1}{1 - ap(0)} \sum_{k=1}^{s} \left(a + \frac{bk}{s}\right) p(k) f(s-k), s = 1{,}2, \ldots \tag{6.3}$$

Důkaz je převzat z (Kaas a kol., 2008) a uvádím ho zde kvůli postupu konstrukce pravděpodobností f(s).

Důkaz. Platí

$$\mathsf{P}(S = 0) = \sum_{n=0}^{\infty} \mathsf{P}(N = n) p^n(0).$$

Tedy pokud $p(0) = 0$, pak:

$$f(0) = \mathsf{P}(N = 0),$$

a pokud $p(0) > 0$, pak:

$$f(0) = \sum_{n=0}^{\infty} \mathsf{P}(N = n) e^{n \log p(0)} = \mathsf{E}\ e^{N \log p(0)} = m_N(\log p(0)).$$

Tím máme dokázanou první část tvrzení (6.3).

Označíme si $S_k = X_1, \ldots, X_k$. Jelikož jsou $X_1, \ldots, X_n$ nezávislé a stejně rozdělené platí:

$$\mathsf{E}\left[a + \frac{bX_1}{s} \middle| S_k = s\right] = a + \frac{b}{k} \tag{6.4}$$

Předchozí rovnost můžeme vyjádřit i jako:

$$\begin{aligned}\mathsf{E}\left[a+\frac{bX_1}{s}\middle|S_k=s\right]&=\sum_{h=0}^{s}\left(a+\frac{bh}{s}\right)\mathsf{P}(X_1=h|S_k=s)\\&=\sum_{h=0}^{s}\left(a+\frac{bh}{s}\right)\frac{\mathsf{P}(X_1=h)\,\mathsf{P}(S_k-X_1=s-h)}{\mathsf{P}(S_k=s)}\end{aligned}\tag{6.5}$$

Z rovností (6.4) a (6.5) plyne:

$$a+\frac{b}{k}=\sum_{h=0}^{s}\left(a+\frac{bh}{s}\right)\frac{\mathsf{P}(X_1=h)\,\mathsf{P}(S_k-X_1=s-h)}{\mathsf{P}(S_k=s)}$$

Z vlastnosti (6.1) a předchozí rovnosti plyne:

$$\begin{aligned}f(s)&=\sum_{k=1}^{\infty}q(k)\,\mathsf{P}(S_k=s)\overset{(6.1)}{=}\sum_{k=1}^{\infty}\left(a+\frac{b}{k}\right)q(k-1)\,\mathsf{P}(S_k=s)\\&\overset{(6.5)}{=}\sum_{k=1}^{\infty}\sum_{h=0}^{s}\left(a+\frac{bh}{s}\right)\mathsf{P}(X_1=h)\,\mathsf{P}(S_k-X_1=s-h)q(k-1)\\&=\sum_{h=0}^{s}\left(a+\frac{bh}{s}\right)\mathsf{P}(X_1=h)\sum_{k=1}^{\infty}\mathsf{P}(S_k-X_1=s-h)q(k-1)\\&=\sum_{h=0}^{s}\left(a+\frac{bh}{s}\right)p(h)f(s-h)=ap(0)f(s)+\sum_{h=1}^{s}\left(a+\frac{bh}{s}\right)p(h)f(s-h)\end{aligned}$$

Úpravou poslední rovnice dokážeme i druhou část tvrzení (6.3).

❑

V Klugman a kol. (2012) můžeme najít Panjerovu formuli pro rozdělení počtu škod, kde (6.1) platí pro $n = 2,3,\ldots$.

6.1 Rozdělení vhodná pro použití Panjerovy rekurze

Mezi rozdělení, které splňují podmínku 6.1, patří pouze:

- Poissonovo rozdělení s parametrem $\lambda \geq 0$, kde $a = 0$ a $b = \lambda$
- Negativně binomické rozdělení s parametry $r > 0$ a $p \in [0,1]$, kde $a = 1 - p$ a $b = a(r-1)$
- Binomické rozdělení s parametry $k \geq 0$ a $p \in [0,1]$, kde $a = -\frac{p}{1-p}$ a $b = -a(k+1)$

.

6.2 Aplikace Panjerovy rekurze

Nyní aplikujeme Panjerovu formuli na vybrané příklady převzaté z (Klugman a kol., 2012) a (Kaas a kol., 2008).

Příklad 3.5.1 z (Kaas a kol., 2008)
Něchť S je složené Poissonovo rozdělení s parametrem $\lambda = 2$ a $p(x) = \mathsf{P}(X = x) = x/10$, $x = 1,2,3,4$. Spočtěte $f(s) = \mathsf{P}(S = s)$ pro $s \leq 4$.

Řešení:
Poissonovo(2) rozdělení patří k rozdělením, na které můžeme aplikovat Panjerovu formuli, kde $a = 0$ a $b = \lambda = 2$.
Jelikož $p(0) = 0$, tak podle (6.2) $f(0) = \mathsf{P}(N = 0) = e^{-2}$. Další hodnoty $f(s)$ vypočteme rekurzivně podle (6.3).

$$
\begin{aligned}
f(1) &= \frac{2}{10}f(0) = \frac{2}{10}e^{-2} \doteq 0.027 \\
f(2) &= \frac{1}{10}f(1) + \frac{4}{10}f(0) = \frac{2}{10^2}e^{-2} + \frac{4}{10}e^{-2} = \frac{42}{10^2}e^{-2} \doteq 0.057 \\
f(3) &= \frac{2}{3\cdot 10}f(2) + \frac{8}{3\cdot 10}f(1) + \frac{6}{10}f(0) = \frac{84}{3\cdot 10^3}e^{-2} + \frac{16}{3\cdot 10^2}e^{-2} + \frac{18}{3\cdot 10}e^{-2} \\
&= \frac{2044}{3\cdot 10^3}e^{-2} \doteq 0.092 \\
f(4) &= \frac{1}{2\cdot 10}f(3) + \frac{2}{10}f(2) + \frac{9}{2\cdot 10}f(1) + \frac{8}{10}f(0) \\
&= \frac{1022}{3\cdot 10^4}e^{-2} + \frac{84}{10^3}e^{-2} + \frac{9}{10^2}e^{-2} + \frac{8}{10}e^{-2} = \frac{30242}{3\cdot 10^4}e^{-2} \doteq 0.136
\end{aligned}
$$

Příklad 3.5.10 z (Kaas a kol., 2008)
Něchť S je složené Poissonovo rozdělení s parametrem $\lambda = 3$ a $p(1) = \frac{5}{6}$, $p(2) = \frac{1}{6}$, kde $p(x) = \mathsf{P}(X = x)$.Spočtěte $f(s) = \mathsf{P}(S = s)$ a $F(s) = P(S \leq s)$, pro $s = 0,1,2,\ldots$.

Řešení:
Poissonovo(3) rozdělení patří k rozdělením, na které můžeme aplikovat Panjerovu formuli, kde $a = 0$ a $b = \lambda = 3$.
Jelikož $p(0) = 0$, tak podle (6.2) $f(0) = \mathsf{P}(N = 0) = e^{-3}$. Rekurzivní vzorec pro další hodnoty $f(s)$ určíme podle (6.3).

$$
f(s) = \frac{1}{s}\left(\frac{15}{6}f(s-1) + f(s-2)\right), \; s = 1,2,\ldots,
$$

kde $f(x) = 0$ pro $x < 0$.
Distribuční funkce potom je $F(s) = \sum_{k=0}^{s} f(k)$.

Příklad 9.44 z (Klugman a kol., 2012)
Složené Poissonovo rozdělení má pět očekávaných škod za rok. Výše škod je kladný násobek tisíce. Označme $f(s) = P(S = s)$ a $p(x) = P(x = x)$. Za podmínek, že

$f(1) = e^{-5}$ a $f(2) = \frac{5}{2}e^{-5}$, spočítejte $p(2)$.

Řešení:
Počet škod má Poissonovo rozdělení s parametrem $\lambda = 5$ a patří k rozdělením, na které můžeme aplikovat Panjerovu formuli, kde $a = 0$ a $b = 5$. Výše škod je kladná náhodná veličina, tedy $p(0) = 0$ a dle (6.2) $f(0) = \mathsf{P}(N = 0) = e^{-5}$. Podle vzorce (6.3) si vyjádříme i $f(1)$ a $f(2)$.

$$f(1) = 5p(1)e^{-5}$$
$$f(2) = \frac{1}{2}(5p(1)f(1) + 10p(2)e^{-5})$$

Dosazením zadaných hodnot plyne z první rovnice $p(1) = \frac{1}{5}$ a z druhé rovnice $p(2) = \frac{2}{5}$.

Příklad 9.52 z (Klugman a kol., 2012)
Pro složené Poissonovo rozdělení, kde výše škod je kladná náhodná veličina, platí

$$f(s) = \frac{1}{x}\left[0.16f(x-1) + kf(x-2) + 0.72f(x-3\right].$$

Očekávaná hodnota celkové škody je 1.68. Spočtěte očekávaný počet škod λ.

Řešení:
Nejdříve si vyjádříme očekávanou výši škod v závislosti na λ. Ze vzorce (6.3) odvodíme, že:

$$p(1) = \frac{0.16}{\lambda}$$
$$p(2) = \frac{k}{2\lambda}$$
$$p(3) = \frac{0.72}{3\lambda}.$$

Výše škod je kladná a tedy $p(0) = 0$. Z rovnosti $p(1) + p(2) + p(3) = 1$ určíme $k = 2\lambda - 0.8$.
Očekávaná výše škod je

$$\mathsf{E}\,X = p(1) + 2p(2) + 3p(3) = \frac{0.08 + 2\lambda}{\lambda}.$$

Nyní využijeme vlastnosti (5.1) a dosadíme.

$$\mathsf{E}\,S = \mathsf{E}[N]\,\mathsf{E}[X]$$
$$1.68 = \lambda\frac{0.08 + 2\lambda}{\lambda}$$
$$\lambda = 0.8$$

Tedy očekávaný počet škod je 0.8.

Příklad 9.57 z (Klugman a kol., 2012)
Celkové škody mají složené Poissonovo rozdělení s parametrem $\lambda = 3$ a rozdělení výše škod $p(1) = 0.4$, $p(2) = 0.3$, $p(3) = 0.2$ a $p(4) = 0.1$, kde $p(x) = \mathsf{P}(X = x)$. Spočtěte pravděpodobnost, že celková škoda nebude vyšší než 3.

Řešení:
Poissonovo rozdělení patří k rozdělením, na které můžeme aplikovat Panjerovu formuli, kde $a = 0$ a $b = \lambda = 3$. Podle vzorce (6.2) platí $f(0) = P(N = 0) = e^{-3}$, protože $p(0) = 0$. Hodnoty $f(1)$, $f(2)$, $f(3)$ získáme rekurzivně z (6.3).

$$\begin{aligned}
f(1) &= 1.2f(0) = 1.2e^{-3} \\
f(2) &= \frac{1}{2}(1.2f(1) + 1.8f(0)) = 0.72e^{-3} + 0.9e^{-3} = 1.62e^{-3} \\
f(3) &= \frac{1}{3}(1.2f(2) + 1.8f(1) + 1.8f(0)) = 0.648e^{-3} + 0.72e^{-3} + 0.6e^{-3} = 1.968e^{-3}.
\end{aligned}$$

Pravděpodobnost, že celková škoda není větší než 3 je:

$$P(S \leq 3) = f(0) + f(1) + f(2) + f(3) = 5.768e^{-3} \doteq 0.287$$

Příklad 9.58 z (Klugman a kol., 2012)
Celkové škody mají složené negativně binomické rozdělení s parametry $r = 6$ $\beta = 0.5$ a rozdělení výše škod $p(1) = 0.4$, $p(2) = 0.3$, $p(3) = 0.2$ a $p(4) = 0.1$, kde $p(x) = \mathsf{P}(X = x)$. Spočtěte pravděpodobnost, že celková škoda nebude vyšší než 3.

Řešení:
Negativně binomické rozdělení patří k rozdělením, na které můžeme aplikovat Panjerovu formuli, kde $a = 1 - \beta = 0.5$ a $b = a(r-1) = 2.5$. Podle vzorce (6.2) platí $f(0) = P(N = 0) = 0.5^6$, protože $p(0) = 0$. Hodnoty $f(1)$, $f(2)$, $f(3)$ získáme rekurzivně z (6.3).

$$\begin{aligned}
f(1) &= (0.5 + 2.5)0.4f(0) = 1.2 \cdot 0.5^6 \\
f(2) &= (0.5 + 1.25)0.4f(1) + (0.5 + 2.5)0.3f(0) = 0.84 \cdot 0.5^6 + 0.9 \cdot 0.5^6 = 1.74 \cdot 0.5^6 \\
f(3) &= \left(0.5 + \frac{5}{6}\right)0.4f(2) + \left(0.5 + \frac{5}{3}\right)0.3f(1) + (0.5 + 2.5)0.2f(0) = 2.308 \cdot 0.5^6.
\end{aligned}$$

Pravděpodobnost, že celková škoda není větší než 3 je:

$$P(S \leq 3) = f(0) + f(1) + f(2) + f(3) = 6.248 \cdot 0.5^6 \doteq 0.098$$

Kapitola 7

Rychlá Fourierova transformace

V této kapitole představíme další z metod výpočtu $f(x) = P(S = s)$ pomocí rychlé Fourierovy transformace. Předpokládejme, že výše škod $X_1, X_2, \dots \sim X$ nabývají pouze hodnot $\{0,1,\dots\}$

Označíme si $\phi_S(t)$ charakteristickou funkci náhodné veličiny S. Pravděpodobnosti $f(s) = P(S = s)$ můžeme vyjádřit pomocí charakteristické funkce následovně:

$$f(s) = \frac{1}{2\pi}\int_0^{2\pi} e^{-its}\phi_S(t)dt. \tag{7.1}$$

Důkaz je uveden například v Kaas a kol. (2008)

Použijeme-li obdélníkové pravidlo na rovnost (7.1) s intervaly délky $2\pi/n$, můžeme aproximovat pravděpodobnosti $f(s)$ jako

$$f(s) \approx 1/n \sum_{k=0}^{n-1} e^{-\frac{isk2\pi}{n}}\phi_S\left(\frac{k2\pi}{n}\right), \tag{7.2}$$

kde chyba je maximálně $\frac{2\pi^3}{3n^2} \max\limits_{t\in(0,2\pi)} \left(\frac{\partial^3}{\partial t^3} e^{-its}\phi_S(t)\right)$.

Stačí počítat pouze s $s = 0, 1, \dots, n-1$, protože například pro $s = n, 2n, \dots$ dostaneme stejnou aproximaci jako v případě $s = 0$.

Nyní představíme diskrétní Fourierovu transformaci vektoru $\vec{f} = (f(0), \dots, f(n-1))$ jako vektor $\vec{y} = \mathbf{T}^-\vec{f}$, kde $\mathbf{T}^-$ je matice $n \times n$ definovaná následovně:

$$\mathbf{T}^-_{jk} = e^{\frac{-i2\pi jk}{n}}, \quad j,k = 0,1,\dots,n-1.$$

Podobně definujeme matici $\mathbf{T}^+$, kde

$$\mathbf{T}^+_{jk} = e^{+\frac{i2\pi jk}{n}}, \quad j,k = 0,1,\dots,n-1.$$

Pomocí aproximace (7.2) a tvaru charakteristické funkce složeného rozdělení (5.3) můžeme vektor $\vec{f}$ aproximovat následovně:

$$\vec{f} \approx r = \frac{1}{n}\mathbf{T}^- P_N(\mathbf{T}^+\vec{p}), \tag{7.3}$$

kde P_N je pravděpodobnostní vytvořující funkce počtu škod N a $\vec{p} = (\mathsf{P}(X = 0), \dots, \mathsf{P}(X = n-1))^T$.

Chyba $r_s - f_s = \mathsf{P}[S \in \{s+n, s+2n, \dots\}]$, $s = 0, \dots, n-1$.

7.1 Aplikace Fourierovy transformace

V této kapitole budeme zjišťovat pravděpodobnosti $f(s) = \mathsf{P}(S = s)$ pomocí rychlé Fourierovy transformace. Jelikož počítáme s velkými maticemi (dimenze $n \times n$), budeme používat matematický software R. Ten má již zabudovanou funkci $\text{fft}(z) = T^{+}z$ a $\text{fft}(z, \text{inverse=True}) = T^{-}z$.

Budeme počítat některé příklady z kapitoly 6.2, kde místo Panjerovy formule budeme využívat rychlou Fourierovu transformaci.

Příklad 3.5.1 z (Kaas a kol., 2008)
Něchť S je složené Poissonovo rozdělení s parametrem $\lambda = 2$ a $p(x) = \mathsf{P}(X = x) = x/10$, $x = 1,2,3,4$. Spočtěte $f(s) = \mathsf{P}(S = s)$ pro $s \leq 4$.

Řešení:
Nejprve musíme zvolit hodnotu n. Tu volíme dostatečně velkou, aby platilo, že $\mathsf{P}(S > n)$ je zanedbatelné. Volím $n = 50$. Podle vzorce (7.3) potřebujeme pravděpodobnostní vytvořující funkci $P_N(t) = e^{2(t-1)}$. Potom:

$$\vec{f} = (f(0),f(1),\ldots,f(49))^T \approx \frac{1}{50}T^{-}e^{2(T^{+}\vec{p}-1)},$$

kde $\vec{p} = (p(0),\, p(1), \ldots,\, p(49))^T = \frac{1}{10}(0,\, 1,\, 2,\, 3,\, 4,\, 0, \ldots,\, 0)^T$. Program R nám vrátí výsledek:

$$\vec{f} = \begin{pmatrix} 0.1353 \\ 0.0271 \\ 0.0568 \\ 0.0922 \\ 0.1364 \\ \vdots \end{pmatrix}.$$

Tedy $f(0) = 0.1353$, $f(1) = 0.0271$, $f(2) = 0.0568$, $f(3) = 0.0922$, $f(4) = 0.1364$.

Příklad 9.57 z (Klugman a kol., 2012)
Celkové škody mají složené Poissonovo rozdělení s parametrem $\lambda = 3$ a rozdělení výše škod $p(1) = 0.4$, $p(2) = 0.3$, $p(3) = 0.2$ a $p(4) = 0.1$, kde $p(x) = \mathsf{P}(X = x)$. Spočtěte pravděpodobnost, že celková škoda nebude vyšší než 3.

Řešení:
Nejprve musíme zvolit hodnotu n. Tu volíme dostatečně velkou, aby platilo, že $\mathsf{P}(S > n)$ je zanedbatelné. Volím $n = 70$. Podle vzorce (7.3) potřebujeme pravděpodobnostní vytvořující funkci $P_N(t) = e^{3(t-1)}$. Potom:

$$\vec{f} = (f(0),\, f(1), \ldots,\, f(69))^T \approx \frac{1}{70}T^{-}e^{3(T^{+}\vec{p}-1)},$$

kde $\vec{p} = (p(0),\, p(1), \ldots,\, p(69))^T = (0,\, 0.4,\, 0.3,\, 0.2,\, 0.1,\, 0, \ldots,\, 0)^T$. Program R *nám vrátí výsledek:*

$$\vec{f} = \begin{pmatrix} 0.0498 \\ 0.0597 \\ 0.0807 \\ 0.098 \\ 0.1075 \\ \vdots \end{pmatrix}.$$

Tedy $f(0) = 0.0498$, $f(1) = 0.0597$, $f(2) = 0.0807$, $f(3) = 0.098$. *Pravděpodobnost, že celková škoda není větší než 3 je:*

$$\mathsf{P}(S \leq 3) = f(0) + f(1) + f(2) + f(3) = 0.2882.$$

Příklad 9.58 z (Klugman a kol., 2012)

Celkové škody mají složené negativně binomické rozdělení s parametry $r = 6$ $\beta = 0.5$ a rozdělení výše škod $p(1) = 0.4$, $p(2) = 0.3$, $p(3) = 0.2$ a $p(4) = 0.1$, kde $p(x) = \mathsf{P}(X = x)$. Spočtěte pravděpodobnost, že celková škoda nebude vyšší než 3.

Řešení:

Nejprve musíme zvolit hodnotu n. Tu volíme dostatečně velkou, aby platilo, že $\mathsf{P}(S > n)$ *je zanedbatelné. Volím* $n = 100$. *Podle vzorce* (7.3) *potřebujeme pravděpodobnostní vytvořující funkci* $P_N(t) = \left(\frac{0.5}{1-0.5t}\right)^6$. *Potom:*

$$\vec{f} = (f(0), f(1), \ldots, f(99))^T \approx \frac{1}{100} T^{-} \left(\frac{0.5}{1 - 0.5T^{+}\vec{p}}\right)^6$$

kde $\vec{p} = (p(0), p(1), \ldots, p(99))^T = (0, 0.4, 0.3, 0.2, 0.1, 0, \ldots, 0)^T$. *Program R nám vrátí výsledek:*

$$\vec{f} = \begin{pmatrix} 0.0156 \\ 0.0188 \\ 0.0272 \\ 0.0361 \\ 0.0441 \\ \vdots \end{pmatrix}.$$

Tedy $f(0) = 0.0156$, $f(1) = 0.0188$, $f(2) = 0.0272$, $f(3) = 0.0361$. *Pravděpodobnost, že celková škoda není větší než 3 je:*

$$\mathsf{P}(S \leq 3) = f(0) + f(1) + f(2) + f(3) = 0.0977.$$

Kapitola 8

Aproximace celkové škody

V této kapitole se budeme zabývat aproximací složeného rozdělení, tedy se budeme snažit přiblížit rozdělení S některým známým rozdělením. Nejčastěji se k aproximaci používají normální, gamma a logaritmicko-normální rozdělení.

8.1 Aproximace normálním rozdělením

Označme si $\mu = \mathsf{E}\ S$ a $\sigma^2 = \operatorname{var} S$. Pro náhodnou veličinu S platí:

$$(S \leq s) = \mathsf{P}\left(\frac{S-\mu}{\sigma} \leq \frac{s-\mu}{\sigma}\right) \doteq \Phi\left(\frac{s-\mu}{\sigma}\right), \tag{8.1}$$

kde $\Phi(x)$ je distribuční funkce normovaného normálního rozdělení.

Například pro složené Poissonovo rozdělení můžeme předchozí tvrzení dokázat pomocí centrální limitní věty.

Věta 3. *(Centrální limitní věta pro složené Poissonovo rozdělení)*
Nechť S má složené Poissonovo rozdělení s parametrem λ a nechť výše škod má rozdělení s konečným rozptylem. Označme $\mu = \mathsf{E}\ S$ a $\sigma^2 = var S$. Potom

$$\lim_{\lambda \to \infty} P\left(\frac{S-\mu}{\sigma} \leq x\right) = \Phi(x),$$

kde $\Phi(x)$ je distribuční funkce normovaného normálního rozdělení.

Důkaz. Jestliže má náhodná veličina Poissonovo rozdělení s celočíselným parametrem λ, můžeme ji zapsat, jako součet λ náhodných veličin s Poissonovým rozdělením s parametrem 1. Tedy:

$$N \sim \sum_{k=1}^{\lambda} N_k, \quad \text{a} \quad S \sim \sum_{k=1}^{\lambda} \sum_{i=1}^{N_k} X_{ik},$$

kde N_k má Poissonovo rozdělení s parametrem 1.

$\sum_{i=1}^{N_k} X_{ik}$ jsou pro všechna $k = 1,2,\ldots,\lambda$ nezávislé a stejně rozdělené náhodné veličiny. Tedy můžeme aplikovat centrální limitní větu a platí dokazované tvrzení. Jestliže λ, není celočíselná tak pro velká λ je necelá část λ zanedbatelná. ❑

8.2 Aproximace logaritmicko-normálním rozdělením

Jestliže náhodná veličina Z má logaritmicko normální rozdělení s parametry μ a σ^2, pak platí $\mathsf{E}\ Z = \exp(\mu + \frac{1}{2}\sigma^2)$ a $\mathsf{E}\,(Z^2) = \exp(2\mu + 2\sigma^2)$.

K aproximaci $S = \sum_{i=1}^{N} X_i$ logaritmicko-normálním rozdělením porovnáme první dva momenty náhodných veličin S a Z. Tedy $\mathsf{E}\ S = \mathsf{E}\ Z$ a $\mathsf{E}\,(S^2) = \operatorname{var} S + (\mathsf{E}\ S)^2 = E(Z^2)$. Z těchto rovnic můžeme vypočítat hodnoty parametrů μ a σ^2. Potom platí *(připomeňme, že náhodná veličina* $\log Z$ *má* $N(\mu,\sigma^2)$ *rozdělení)*:

$$\begin{aligned} \mathsf{P}(S \leq s) &\doteq \mathsf{P}(Z \leq s) = \mathsf{P}(\log Z \leq \log s) \\ &= \mathsf{P}\left(\frac{\log Z - \mu}{\sigma} \leq \frac{\log s - \mu}{\sigma}\right) = \Phi\left(\frac{\log s - \mu}{\sigma}\right), \end{aligned} \tag{8.2}$$

kde $\Phi(x)$ je distribuční funkce normálního rozdělení.

Aproximace normálním rozdělením dává dobré výsledky pokud $\mathsf{E}\ S$ je velké. Například pokud N má Poissonovo(λ), binomické(m,p) nebo negativně binomické(r,p) rozdělení, tak s využitím centrální limitní věty posíláme parametry λ, m a r do nekonečna. *Pro Poissonovo rozdělení plyne z věty 3, pro binomické a negativně binomické dokážeme podobně.* Pokud $\mathsf{E}\,[S]$ je malé, potom je rozdělení S zešikmené a tedy vhodnější aproximací je logaritmicko-normální rozdělení.

Další možné aproximace můžeme najít třeba v Mandl a Mazurová (1999).

8.3 Příklad

V této části si ukážeme příklad na aproximaci celkové škody. Příklad je převzat z (Kaas a kol., 2008).

Příklad 3.7.1

Nechť S je složené Poissonovo rozdělení s parametrem $\lambda = 12$ a rovnoměrným rozdělením R(0,1) výše škod. Aproximujte $\mathsf{P}(S \leq 10)$ pomocí normálního a logaritmicko-normálního rozdělení.

Řešení:

Nejdříve potřebujeme určit střední hodnotu a rozptyl složeného rozdělení. tomu využijeme rovnosti (5.1) *a* (5.2)*:*

$$E[S] = E[X]\,E[N] = 12 \cdot 0.5 = 6,$$

$$var[S] = E[N]\,var[X] + (E[X])^2\,var[N] = 12 \cdot \frac{1}{12} + (0.5)^2 12 = 4$$

Normální rozdělení:

Podle rovnosti (8.1) *platí:*

$$P(S \leq 10) \doteq \Phi\left(\frac{10 - E\ S}{\sqrt{var\,S}}\right) = \Phi(2) = 0.9772$$

Logaritmicko-normální rozdělení:
Nejdříve musíme zjistit parametry μ *a* σ^2 *logaritmicko-normálního rozdělení porovnáním prvních dvou momentů:*

$$6 = \mathsf{E}[S] \stackrel{!}{=} e^{\mu+\frac{1}{2}\sigma^2}$$

$$40 = \mathsf{E}[S^2] \stackrel{!}{=} e^{2\mu+2\sigma^2}.$$

Vyřešením soustavy získáme $\mu = \frac{1}{2}\log(\frac{162}{5})$ *a* $\sigma^2 = \log(\frac{10}{9})$.
Podle rovnosti (8.2) *platí:*

$$\mathsf{P}(S \leq 10) = \Phi\left(\frac{\log 10 - \mu}{\sigma}\right) \doteq \Phi(1.736) = 0.9587$$

Kapitola 9

Závěr

Tato práce byla zaměřena na výpočty složeného rozdělení kolektivního modelu. Představila jsem nejčastěji využívané rozdělení pro modelování počtu a výše škod. Metodou maximální věrohodnosti jsme u každého ze zmíněných rozdělení určili odhady parametrů.

Další kapitoly jsme již věnovali složenému rozdělení. Ukázali jsme jeho základní vlastnosti, které můžeme využít k určení rozdělení. Představili jsme dva způsoby výpočtu rozdělení agregovaných škod a to Panjerovu rekurzi a rychlou Fourierovu transformaci. U obou možností je nevýhodou, že používáme pouze diskrétní rozdělení výše škod. Abychom mohli určit rozdělení agregovaných škod se spojitým rozdělením výše škod již zmíněnými metodami, museli bychom rozšířit teorii například o diskretizaci náhodných veličin.

V poslední kapitole se věnuji aproximaci agregovaných škod pomocí normálního a logaritmicko-normálního rozdělení. Tuto metodu využíváme pro nejjednodušší výpočet distribuční funkce složeného rozdělení. Existuje ještě mnoho způsobů aproximace, kterými by se dalo pokračovat.

Práce obsahuje i řešené příklady na Panjerovu rekurzi, rychlou Fourierovu transformaci a aproximaci složeného rozdělení. Příklady jsou vybrány tak, aby nebyly složité a dala se na ně jednoduše aplikovat získaná teorie.

Seznam použité literatury

KAAS, R., GOOVAERTS, M. AMD DHAENE, J. a DENUIT, M. (2008). *Modern Actuarial Risk Theory Using R.* 2nd edition. Springer. ISBN 978-3-540-70992-3.

KLUGMAN, S., PANJER, H. a WILLMONT (2012). *Loss Models: From Data to Decisions.* 4th edition. John Wiley & Sons. ISBN 978-1-118-31532-3.

KOUDELKA, J. (2014). Modely kolektivního rizika v neživotním pojištění. Bakalářská práce, Přírodovědecká fakulta, Masarikova univerzita, Ústav matematiky a statistiky. Vedoucí bakalářské práce Mgr. Jan Koláček, Ph.D.

MANDL, P. a MAZUROVÁ, L. (1999). *Matematické základy neživotního pojištění.* Matfyzpress. ISBN 80-85863-42-1.