Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Katarína Valentovičová

Složené Poissonovo rozdělení

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce:	RNDr. Šárka Hudecová, Ph.D.
Studijní program:	Matematika
Studijní obor:	Finanční matematika

Praha 2013

Na tomto mieste by som sa rada poďakovala vedúcej mojej bakalárskej práce, RNDr. Šárke Hudecovej, Ph.D., za odborné vedenie, užitočné rady a pripomienky a za veľkú trpezlivosť a ústretovosť pri jej spracovaní.
Srdečné poďakovanie patrí taktiež mojim rodičom za podporu počas priebehu štúdia.

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) 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

Název práce: Složené Poissonovo rozdělení

Autor: Katarína Valentovičová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Šárka Hudecová, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Stanovení rezerv a odhad předpokládaného škodního průběhu jsou jedním ze základních problémů pojišťovnictví, ve kterých se uplatňují statistické metody. Jednou z nich je modelování s využitím složeného rozdělení. Složené rozdělení vzniká jako součet náhodného počtu nezávislých stejně rozdělených náhodných veličin. Tato práce se zabývá výkladem složeného Poissonova rozdělení, jeho vlastnostmi a využitím v praxi, přičemž se soustředí na jeho aplikaci v neživotním pojištění. Uvádí postup odvození hustoty takto zkonstruovaného rozdělení a pojednává o metodách odhadu parametrů. V závěru práce aplikujeme teoretické vědomosti na reálná data.

Klíčová slova: složené rozdělení, Tweedieho složený Poissonův model, odhady parametrů, rezerva na pojistná plnění

Title: Compound Poisson distribution

Author: Katarína Valentovičová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Šárka Hudecová, Ph.D., Department of Probability and Mathematical Statistics

Abstract: Claims reserving and claims process estimation are classical problems in general insurance. Some of the statistical methods in this field are based on a compound distribution. This distribution arises as a sum of a random number of independent and identically distributed variables. This thesis deals, in particular, with the compound Poisson distribution, its properties and possible applications in general insurance. Basic theoretical properties of the distribution are derived, and parameters estimation methods are discussed. The theoretical methods are illustrated on a real data set from car insurance.

Keywords: compound distribution, Tweedie´s compound Poisson model, parameters estimations, claims reserving

Názov práce: Zložené Poissonovo rozdelenie

Autor: Katarína Valentovičová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: RNDr. Šárka Hudecová, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Stanovenie rezerv a odhad predpokladaného škodného priebehu sú jedným zo základných problémov poisťovníctva, v ktorých sa uplatňujú štatistické metódy. Jednou z nich je modelovanie využitím zloženého rozdelenia. Zložené rozdelenie vzniká ako súčet náhodného počtu nezávislých rovnako rozdelených náhodných veličín. Táto práca sa zaoberá výkladom zloženého Poissonovho rozdelenia, jeho vlastností a využitia v praxi, pričom sa sústredí na jeho aplikáciu v neživotnom poistení. Uvádza postup odvodenia hustoty takto skonštruovaného rozdelenia za predpokladu konkrétne zvolených pravdepodobnostných rozdelení a pojednáva taktiež o metódach odhadu parametrov. V závere práce aplikujeme teoretické vedomosti na reálne dáta.

Kľúčové slová: zložené rozdelenie, Tweedieho zložený Poissonov model, odhady parametrov, rezerva na poistné plnenia

Obsah

Úvod

Poisťovníctvo ako jedno z významných odvetví národného hospodárstva poskytuje na trhu špecifický tovar – poistenie. Hlavným cieľom poistenia je náhrada škôd vzniknutých pri náhodných udalostiach. Pri riešení významných otázok svojej činnosti poisťovňa vyžaduje určité poznanie, ktoré sa týka frekvencie výskytu poistných udalostí a taktiež výšky poistných plnení v prípade ich vzniku. V jazyku teórie pravdepodobnosti potrebuje poisťovňa poznať rozdelenie počtu poistných plnení a rozdelenie výšky poistných plnení, či už individuálnych alebo celkových.

Pri matematickej analýze činnosti poisťovní je jedným zo základných problémov modelovanie výšky škôd, ktoré nastali pre všetky poistné zmluvy v určitom časovom horizonte. Uvažujme jednoduchý príklad: predpokladajme, že pravdepodobnosť výskytu poistnej udalosti pri danom type poistenia je $0,07$, t.j. v priemere pri siedmich zo 100 uzavretých zmlúv nastane poistná udalosť, z ktorej plynie poisťovni povinnosť uhradiť odpovedajúce poistné plnenie. Poisťovňa teda potrebuje mať k dispozícii dostatok finančných prostriedkov na vyplatenie týchto siedmich poistných plnení. Princíp je taký, že stanoví poistné tak, aby prijaté poistné zo 100 zmlúv pokrývalo práve daných sedem poistných plnení. Tento opis je veľmi zjednodušený, ale ilustruje, ktoré údaje sú pre poisťovňu dôležité.

Týmto problémom sa zaoberá tzv. model kolektívneho rizika. Obvykle používa nasledovnú matematickú formuláciu: N označuje počet nárokov v uvažovanom období, nech X_1 značí výšku prvého nároku, X_2 druhého nároku atď. a X_N je výška posledného nároku, ktorý bol nahlásený v spomínanom období. Celkovú výšku škôd reprezentuje náhodná veličina S, ktorá je určená súčtom $S = \sum_{i=1}^{N} X_i$, pričom predpokladáme, že výšky jednotlivých nárokov sú nezávislé a tiež že počet nárokov je nezávislý na výške jednotlivých nárokov. Takto skonštruovaná náhodná veličina má zložené rozdelenie, ktoré umožňuje modelovať tvorbu poistných rezerv.

Pre výber témy bakalárskej práce som sa rozhodla na základe progresívneho rozvoja poisťovníctva na trhu a jeho nezastupiteľného miesta v hospodárskom odvetví.

Cieľom predkladanej práce je popísať základné vlastnosti zloženého Poissonovho rozdelenia, metódy odhadu parametrov a uviesť jeho použitie v praxi, pričom sa zameriame najmä na aplikáciu v neživotnom poistení. Aktuálne sa tejto téme vo svojich publikáciách venujú Jørgensen a Paes de Souza (1994), Smyth a Jørgensen (2002) a Wüthrich (2003).

V kapitole 1 spracujeme základnú terminológiu štatistických pojmov, ktoré používame v obsahu predmetnej práce.

V druhej kapitole zavedieme pojem zloženého rozdelenia, popíšeme jeho vlastnosti a zadefinujeme v nej zložené Poissonovo rozdelenie.

V kapitole 3, ktorá tvorí nosnú časť predkladanej práce, uvažujeme konkrétne zvolené rozdelenia, na základe ktorých odvodzujeme charakteristiky Tweedieho zloženého Poissonovho rozdelenia, popisujeme postup odvodenia hustoty a odhadu parametrov.

Vyššie uvedené teoretické poznatky sú následne ilustrované na konkrétnych dátach z oblasti poistenia motorových vozidiel. Príslušnú analýzu dát dokumentuje kapitola 4.

Bakalárska práca je svojím obsahom prínosná ako pre študentov Matematicko-fyzikálnej fakulty Univerzity Karlovej, tak i pre odbornú verejnosť v oblasti poisťovníctva. Za najväčší prínos práce považujeme podrobne vysvetlené postupy odvodenia hustoty a odhadu prametrov. Ako ďalší prínos hodnotíme spracovanie a analýzu dát uvedenú v kapitole 4.

Kapitola 1

Základné pojmy

V prvej kapitole uvedieme definície používaných pojmov, vety a tvrdenia, ktoré sme čerpali najmä z Anděl (2007), Dupač a Hušková (2001) a Zvára a Štěpán (2001). V časti 1.1 pripomenieme pojem náhodnej veličiny a uvedieme niektoré jej charakteristiky a vzťahy medzi nimi. Základné vlastnosti Poissonovho a gama rozdelenia, na ktoré sa budeme v práci odvolávať, popíšeme v podkapitolách 1.2 a 1.3. Časť 1.4 je venovaná podmienenému rozdeleniu, strednej hodnote a rozptylu a v poslednej časti 1.5 uvedieme užitočné vlastnosti súčtu náhodných veličín, ktoré využijeme v druhej kapitole.

1.1 Náhodná veličina a jej vlastnosti

Označme Ω množinu všetkých elementárnych javov. Množinu Ω nazývame priestor elementárnych javov. Nech je na priestore Ω daná nejaká σ-algebra $\mathcal{A}$ jeho podmnožín. Tieto podmnožiny sa nazývajú náhodné javy. Jednotlivým množinám patriacim do $\mathcal{A}$ sa potom pripisuje pravdepodobnosť pomocou nejakej pravdepodobnostnej miery P. Trojica $(\Omega, \mathcal{A}, \mathsf{P})$ sa nazýva pravdepodobnostný priestor.

Nech $\mathbb{R}$ je reálna priamka a $\mathcal{B}$ systém jej borelovských množín. Nech $X(\omega)$ je meratelná funkcia z $(\Omega, \mathcal{A}, \mathsf{P})$ do $(\mathbb{R}, \mathcal{B})$. Potom sa $X(\omega)$ nazýva náhodná veličina resp. náhodná premenná a značí sa stručne X.

Definícia 1.1.1. *Nech Y je náhodná veličina. Komplexná funkcia reálnej premennej $\psi_Y(t) : \mathbb{R} \to \mathbb{C}$, ktorú definujeme nasledujúcim vzťahom:*

$$\psi_Y(t) = \mathsf{E}\left(e^{itY}\right),$$

sa nazýva charakteristická funkcia náhodnej veličiny Y.

Definícia 1.1.2. *Nech Y je náhodná veličina. Reálna funkcia reálnej premennej $M(t) : \mathbb{R} \to [0; \infty]$, ktorú definujeme nasledujúcim vzťahom:*

$$M_Y(t) = \mathsf{E}\left(e^{tY}\right),$$

sa nazýva momentová vytvárajúca funkcia náhodnej veličiny Y.

Vzhľadom na to, ako je momentová vytvárajúca funkcia definovaná, nemusí vždy existovať pre každé $t \in \mathbb{R}$. To je zásadný rozdiel medzi touto a charakteristickou funkciou, ktorá existuje vždy.

Poznámka 1.1.1. Pre spojité náhodné veličiny s hustotou $f(y)$ teda podľa definície strednej hodnoty platí

$$M_Y(t) = \int_{-\infty}^{\infty} e^{ty} f(y)\, dy. \tag{1.1}$$

Pre diskrétne náhodné veličiny dostávame

$$M_Y(t) = \sum e^{ty_i} p_i, \tag{1.2}$$

kde p_i je pravdepodobnostné rozdelenie náhodnej veličiny Y, t.j. $p_i = \mathsf{P}(Y = y_i)$.

Veta 1.1.1. *Nech $M_Y(t) = \mathsf{E}(e^{tY}) < \infty$, pre všetky $|t| \leq t_0$ a nejaké $t_0 > 0$. Potom $\mathsf{E}|Y|^r < \infty$ pre všetky $k > 0$ a platí*

$$\frac{\partial^r M_Y}{\partial t^r}(t)\Big|_{t=0} = \mathsf{E}Y^r,\ r = 1, 2, ... \tag{1.3}$$

Dôkaz. viď Dupač a Hušková (2001), Veta 2.12. na strane 27. □

Rovnosť (1.3) odôvodňuje slovné označenie funkcie $M_Y(t)$. Z hodnoty $M_Y(t)$ môžeme derivovaním jednoducho spočítať všetky momenty.

1.2 Poissonovo rozdelenie

Poissonovo rozdelenie je v teórii pravdepodobnosti a štatistike diskrétne rozdelenie pravdepodobnosti, ktoré môžeme interpretovať ako výskyt zriedkavých udalostí v sérii veľkého počtu nezávislých pokusov. Pomocou tohto rozdelenia sa modeluje výskyt nejakých udalostí, ktoré nastávajú náhodne, napr. počet poistných udalostí v danom časovom úseku. Parameter Poissonovho rozdelenia λ sa často nazýva intenzita výskytu udalostí.

Poissonovo rozdelenie sa často používa ako aproximácia binomického rozdelenia pre veľký počet pokusov, teda pre $n \to \infty$, a malú pravdepodobnosť výskytu sledovaného javu v jednom pokuse, teda $p \to 0$, kde $pn \to \lambda$ (viď Zvára a Štěpán (2001) na strane 66).

Hovoríme, že náhodná veličina Y má Poissonovo rozdelenie s parametrom $\lambda > 0$ (značíme $Y \sim Po(\lambda)$), keď nadobúda hodnoty $j = 0, 1, 2, ...$ s pravdepodobnosťami $\frac{\lambda^j}{j!} e^{-\lambda}, 0 \leq j < \infty$, t.j.

$$\mathsf{P}(Y = \ j) = \frac{\lambda^j}{j!} e^{-\lambda}, \text{ pre } j = 0, 1, 2, ...$$

Príklady rozdelenia pravdepodobností Poissonovho rozdelenia sú uvedené na Obr. 1.1.

Stredná hodnota a rozptyl Poissonovho rozdelenia sú $\mathsf{E}Y = \lambda, \mathsf{var}Y = \lambda$. Tieto hodnoty dostaneme štandardným výpočtom podľa definície strednej hodnoty a rozptylu(viď Dupač a Hušková (2001)).

Spočítajme momentovú vytvárajúcu funkciu Poissonovho rozdelenia. Pre všetky reálne t platí $M_Y(t) = \mathsf{E}(e^{tY})$. Využitím vzťahu (1.2) a rozvoja exponenciály do mocninnej rady dostávame

$$M_Y(t) = \sum_{j=0}^{\infty} e^{tj} \frac{\lambda^j}{j!} e^{-\lambda} = e^{-\lambda} \sum_{j=0}^{\infty} e^{tj} \frac{\lambda^j}{j!} = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(e^t \lambda)^j}{j!} = e^{-\lambda} e^{e^t \lambda}$$
$$= e^{\lambda(e^t - 1)}.$$

Podobne spočítame charakteristickú funkciu $\psi_Y(t)$:

$$\psi_Y(t) = \mathsf{E}(e^{itY}) = \sum_{j=0}^{\infty} e^{itj} \frac{\lambda^j}{j!} e^{-\lambda} = e^{-\lambda} \sum_{j=0}^{\infty} e^{itj} \frac{\lambda^j}{j!} = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(e^{it} \lambda)^j}{j!}$$
$$= e^{-\lambda} e^{e^{it} \lambda} = e^{\lambda(e^{it} - 1)}.$$

1.3 Gama rozdelenie

Gama rozdelenie je v teórii pravdepodobnosti a štatistike spojité rozdelenie pravdepodobnosti s dvoma parametrami.

Hovoríme, že náhodná veličiná Y má gama rozdelenie s parametrami $\alpha > 0$, $\beta > 0$ (značíme $Y \sim Gama(\alpha, \beta)$), keď jej hustota pravdepodobnosti má tvar

$$f_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \mathbb{I}_{(0,\infty)}(y), y \in \mathbb{R},$$

kde

$$\mathbb{I}_{(0,\infty)}(y) = \begin{cases} 1, & \text{ak } y \in (0, \infty), \\ 0, & \text{inak.} \end{cases}$$

Symbolom $\Gamma(\cdot)$ v hustote rozumieme gama funkciu, ktorá je definovaná ako

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} \, dx, \text{ pre } p > 0.$$

Pre gama funkciu platí $\Gamma(p + 1) = p\Gamma(p)$, t.j. $\Gamma(n) = (n - 1)!$ pre $n \in \mathbb{N}$. Naviac platí, že $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Paramater α nazývame parameter tvaru (anglicky *shape* paramater) a β zase parameter škály (anglicky *rate* parameter).

Poznámka 1.3.1. V rôznych literatúrach sa vyskytuje alternatívna parametrizácia gama rozdelenia pomocou *shape* parametra α, ktorý zostáva nezmenený a pomocou parametra mierky γ (anglicky *scale* parameter), pre ktorý platí $\gamma = \frac{1}{\beta}$.

Príklady hustoty gama rozdelenia pre rôzne zvolené parametre sú uvedené na Obr. 1.2.

Stredná hodnota a rozptyl gama rozdelenia sú $\mathsf{E}Y = \dfrac{\alpha}{\beta}, \mathsf{var}Y = \dfrac{\alpha}{\beta^2}$.

Výpočet:

$$\begin{aligned}
\mathsf{E}Y &= \int_0^\infty y \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}\, dy = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^\alpha e^{-\beta y}\, dy \overset{(1)}{=} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\tfrac{s}{\beta}\right)^\alpha e^{-s} \frac{1}{\beta}\, ds \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\beta^{\alpha+1}} \int_0^\infty s^\alpha e^{-s}\, ds = \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)} = \frac{\alpha}{\beta}, \\
\mathsf{E}Y^2 &= \int_0^\infty y^2 \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}\, dy = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha+1} e^{-\beta y}\, dy \overset{(1)}{=} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{s}{\beta}\right)^{\alpha+1} e^{-s} \frac{1}{\beta}\, ds \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\beta^{\alpha+2}} \int_0^\infty s^{\alpha+1} e^{-s}\, ds = \frac{\Gamma(\alpha+2)}{\beta^2\Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{\beta^2}.
\end{aligned}$$

V oboch výpočtoch sme v rovnosti (1) použili substitúciu $s = \beta y$. Dosadením do definície dostávame rozptyl

$$\mathsf{var}Y = \mathsf{E}Y^2 - (\mathsf{E}Y)^2 = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}.$$

Spočítajme momentovú vytvárajúcu funkciu gama rozdelenia. Pre všetky reálne t platí $M_Y(t) = \mathsf{E}(e^{tY})$. Dosadíme do vzťahu (1.1) a počítame

$$M_Y(t) = \int_0^\infty e^{ty} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}\, dy = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-y(\beta-t)} y^{\alpha-1}\, dy.$$

Tento integrál existuje za podmienky $t < \beta$. Použijeme substitúciu $s = y(\beta - t)$ a dostaneme

$$\begin{aligned}
M_Y(t) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-s} \left(\frac{s}{\beta-t}\right)^{\alpha-1} ds = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\beta-t)^\alpha} \int_0^\infty e^{-s} s^{\alpha-1}\, ds \\
&= \frac{\beta^\alpha}{(\beta-t)^\alpha} = \frac{\beta^\alpha}{\beta^\alpha(1-\frac{t}{\beta})^\alpha} = \left(1 - \frac{t}{\beta}\right)^{-\alpha}.
\end{aligned}$$

Podobne by sme spočítali charakteristickú funkciu, ktorá vychádza

$$\psi_Y(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha}.$$

Uvedieme ešte jednu užitočnú vlastnosť gama rozdelenia. Nech X_i má gama rozdelenie s parametrami α_i a β a nech X_i sú nezávislé pre $i = 1, 2, \ldots, n$. Potom náhodná veličina $X = \sum_{i=1}^n X_i$ má tiež gama rozdelenie s paramatrami $\sum_{i=1}^n \alpha_i$ a β. Túto vlastnosť by sme dokazovali matematickou indkuciou, ktorej základom by bola veta o konvolúcii (viď Věta 8.1 v Zvára a Štěpán (2001) na strane 117).

1.4 Podmienené rozdelenie

V tejto kapitole budeme uvažovať náhodné veličiny a náhodné vektory definované na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, \mathsf{P})$. Na úvod pripomenieme definíciu náhodného vektora.

Definícia 1.4.1. *Merateľné zobrazenie* $\boldsymbol{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathcal{B}^n)$, *sa nazýva (n-rozmerný) náhodný vektor.*

Uvažujme náhodný vektor $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$, ktorý je rozdelený na dva podvektory $\boldsymbol{Y} = (X_1, X_2, \ldots, X_r)^T$, a $\boldsymbol{Z} = (X_{r+1}, X_{r+2}, \ldots, X_n)^T, 1 \leq r < n$. Budeme pozorovať rozdelenie náhodného vektoru $\boldsymbol{Y}$ v situácii, že náhodný vektor $\boldsymbol{Z}$ realizoval konkrétnu hodnotu $\boldsymbol{z} \in \mathbb{R}^{n-r}$.

Definícia 1.4.2. *Nech náhodný vektor* $\boldsymbol{Y}$ *má hustotu* $f_{\boldsymbol{Y}}(\boldsymbol{y})$ *vzhľadom k* σ*-konečnej miere* μ_1 *na* $(\mathbb{R}^r, \mathcal{B}^r)$ *a náhodný vektor* $\boldsymbol{Z}$ *má hustotu* $f_{\boldsymbol{Z}}(\boldsymbol{z})$ *vzhľadom k* σ*-konečnej miere* μ_2 *na* $(\mathbb{R}^{n-r}, \mathcal{B}^{n-r})$*. Nech náhodný vektor* $\boldsymbol{X} = (\boldsymbol{Y}^T, \boldsymbol{Z}^T)^T$ *má hustotu* $f_{\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{z})$ *vzhľadom k súčinovej miere* $\mu = \mu_1 \times \mu_2$ *na* $(\mathbb{R}^n, \mathcal{B}^n)$*. Podmienenú hustotu náhodného vektoru* $\boldsymbol{Y}$ *pri danom* $\boldsymbol{Z}$ *nazveme takú nezápornú merateľnú funkciu* $f(\boldsymbol{y} \mid \boldsymbol{z})$*, ktorá pre ľubovolné množiny* $B \in \mathcal{B}^r$ *a* $C \in \mathcal{B}^{n-r}$ *splňuje vzťah*

$$\mathsf{P}(\boldsymbol{Y} \in B, \boldsymbol{Z} \in C) = \int_C \left[\int_B f(\boldsymbol{y} \mid \boldsymbol{z})\, d\mu_1(\boldsymbol{y}) \right] f_{\boldsymbol{Z}}(\boldsymbol{z})\, d\mu_2(\boldsymbol{z}).$$

Stále uvažujeme náhodný vektor $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$, ktorý znovu rozdelíme na dva vektory $\boldsymbol{Y} = (X_1, X_2, \ldots, X_r)^T$, a $\boldsymbol{Z} = (X_{r+1}, X_{r+2}, \ldots, X_n)^T$. Využitím podmienej hustoty zadefinujeme pojem podmienenej strednej hodnoty.

Definícia 1.4.3. *Nech* $g : \mathbb{R}^n \to \mathbb{R}^m$ *je merateľná funkcia, takže* $Q = g(\boldsymbol{X}) = g(\boldsymbol{Y}, \boldsymbol{Z})$ *je náhodný vektor. Nech existuje konečná stredná hodnota* $\mathsf{E}Q$*. Podmienenú strednú hodnotu náhodného vektoru* Q *za podmienky, že vektor* $\boldsymbol{Z}$ *nadobudol hodnoty* $\boldsymbol{z}$*, definujeme vzorcom*

$$\mathsf{E}(Q \mid \boldsymbol{Z} = \boldsymbol{z}) = \int_{\mathbb{R}^r} g(\boldsymbol{y}, \boldsymbol{z}) f(\boldsymbol{y} \mid \boldsymbol{z})\, d\mu_1(\boldsymbol{y}),$$

ak integrál na pravej strane existuje.

Veta 1.4.1. *Nech* $g_1, g_2 : \mathbb{R}^n \to \mathbb{R}^m$ *a* $h : \mathbb{R}^{n-r} \to \mathbb{R}$ *sú merateľné funkcie. Označme* $Q_1 = g_1(\boldsymbol{Y}, \boldsymbol{Z})$ *a* $Q_2 = g_2(\boldsymbol{Y}, \boldsymbol{Z})$*. Nech majú vektory* Q, Q_1 *a* Q_2 *konečné prvé momenty. Potom platí*

1. $\mathsf{E}(a \mid \boldsymbol{Z}) = a$ pre každé $a \in \mathbb{R}^m$,
2. $\mathsf{E}[\mathsf{E}(Q \mid \boldsymbol{Z})] = \mathsf{E}Q$, (1.4)
3. $\mathsf{E}(a_1 Q_1 + a_2 Q_2 \mid \boldsymbol{Z}) = a_1 \mathsf{E}(Q_1 \mid \boldsymbol{Z}) + a_2 \mathsf{E}(Q_2 \mid \boldsymbol{Z})$ pre každé $a_1, a_2 \in \mathbb{R}^m$,
4. $\mathsf{E}(h(\boldsymbol{Z}) Q \mid \boldsymbol{Z}) = h(\boldsymbol{Z}) \mathsf{E}(Q \mid \boldsymbol{Z})$.

Dôkaz. Viď Veta 3.22, 3.23, 3.24 a 3.25 v Anděl (2007) na strane 35. □

Podobne ako podmienenú strednú hodnotu môžeme definovať i podmienený rozptyl. Symbolom $^{\otimes 2}$ v nasledujúcej definícii budeme rozumieť vektorové umocnenie na druhú, t.j. pre vektor $\mathbf{a} \in \mathbb{R}^n$ je $\mathbf{a}^{\otimes 2} = \mathbf{a}\,\mathbf{a}^T$. Nerovnosť $\mathsf{E}Q^{\otimes 2} < \infty$ bude znamenať, že matica $\mathsf{E}Q^{\otimes 2}$ je konečná vo všetkých zložkách.

Definícia 1.4.4. *Nech platia označenia predchádzajúcich definícií a nech* $\mathsf{E}Q^{\otimes 2} < \infty$. *Potom podmienený rozptyl je definovaný rovnicou*

$$\mathsf{var}(Q \mid \boldsymbol{Z}) = \mathsf{E}\{[Q - \mathsf{E}(Q \mid \boldsymbol{Z})]^{\otimes 2} \mid \boldsymbol{Z}\}.$$

Veta 1.4.2. *Ak je* $\mathsf{E}Q^{\otimes 2} < \infty$, *potom platí*

$$\mathsf{var}Q = \mathsf{var}[\mathsf{E}(Q \mid \boldsymbol{Z})] + \mathsf{E}\,\mathsf{var}(Q \mid \boldsymbol{Z}). \tag{1.5}$$

Dôkaz. Viď Veta 3.26 v Anděl (2007) na strane 36. □

1.5 Vlastnosti súčtu náhodných veličín

Uvažujme náhodné veličiny $X_1, ..., X_n, n \in \mathbb{N}$, ktoré sú nezávislé. Označme S náhodnú veličinu, ktorá vznikne súčtom náhodných veličín $X_1, ..., X_n$, t.j. $S = X_1 + \cdots + X_n = \sum_{i=1}^n X_i$. Budeme skúmať momentovú vytvárajúcu funkciu náhodnej veličiny S, ak ja daná momentová vytvárajúca funkcia $M_{X_i}(t)$ náhodnej veličiny $X_i, i = 1, 2, ..., n$.

Veta 1.5.1. *Nech* $X_1, ..., X_n$ *sú nezávislé náhodné veličiny,* $M_{X_i}(t)$ *je momentová vytvárajúca funkcia náhodnej veličiny* $X_i, i = 1, 2, ..., n$. *Potom náhodná veličina* $S = \sum_{i=1}^n X_i$ *má momentovú vytvárajúcu funkciu*

$$M_S(t) = \prod_{i=1}^n M_{X_i}(t).$$

Dôkaz. Počítajme $M_S(t)$:

$$M_S(t) = \mathsf{E}(e^{tS}) = \mathsf{E}\big(e^{t\sum_{i=1}^n X_i}\big) = \mathsf{E}\big(e^{\sum_{i=1}^n tX_i}\big) = \mathsf{E}\Big(\prod_{i=1}^n e^{tX_i}\Big).$$

Použijeme vzťah $\mathsf{E}XY = \mathsf{E}X\mathsf{E}Y$ pre nezávislé náhodné veličiny, v našom prípade pre n nezávislých náhodných veličín $\mathsf{E}\left(\prod_{i=1}^n e^{tX_i}\right) = \prod_{i=1}^n \mathsf{E}\left(e^{tX_i}\right)$:

$$M_S(t) = \prod_{i=1}^n \mathsf{E}\left(e^{tX_i}\right) = \prod_{i=1}^n M_{X_i}(t).$$

□

Nasledujúca veta je dôsledkom vety 1.5.1.

Veta 1.5.2. *Nech* $X_1, ..., X_n$ *sú nezávislé rovnako rozdelené náhodné veličiny s momentovou vytvárajúcou funkciou* $M_X(t)$. *Potom náhodná veličina* $S = \sum_{i=1}^n X_i$ *má momentovú vytvárajúcu funkciu*

$$M_S(t) = (M_X(t))^n.$$

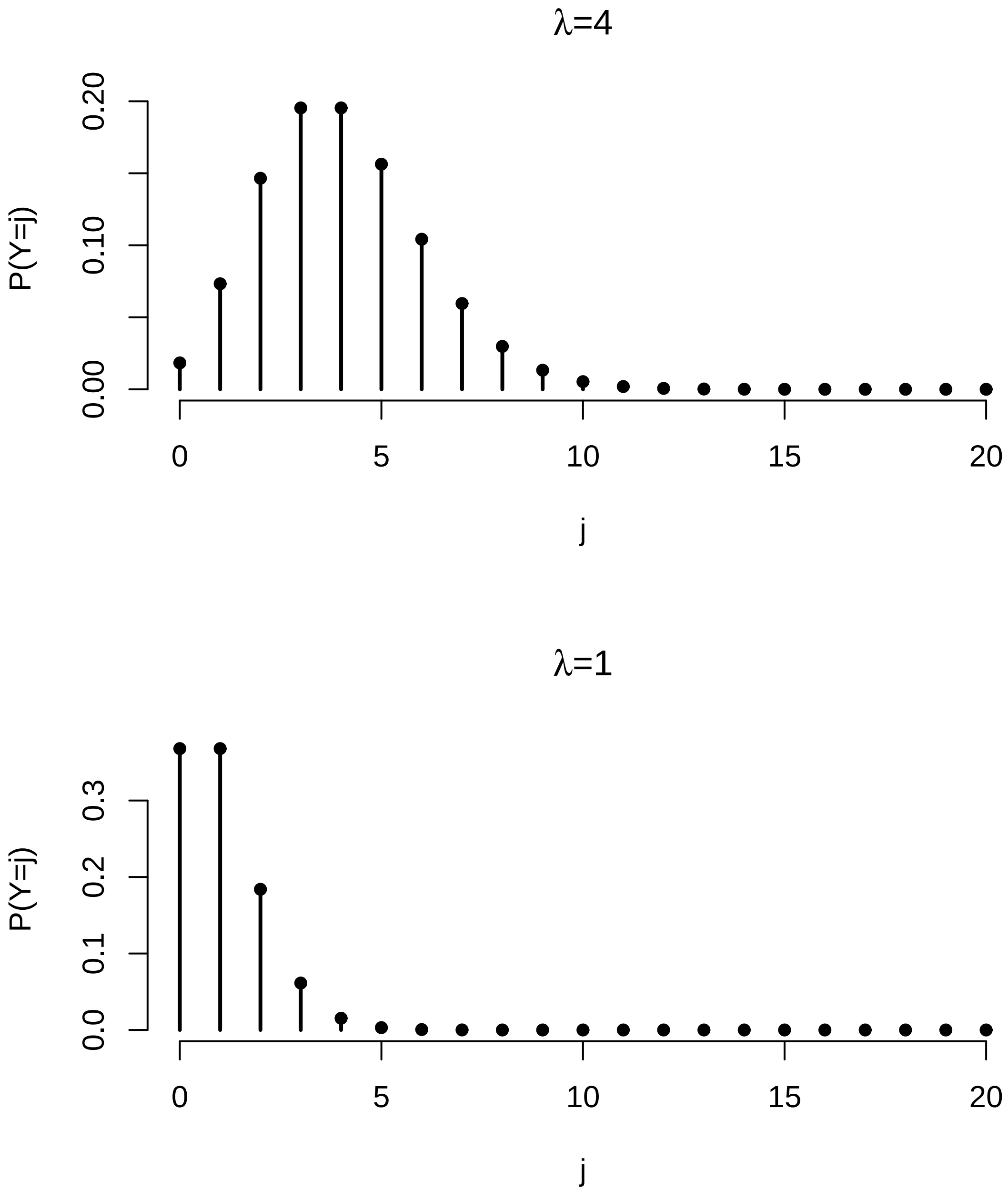
Dôkaz. Podľa vety 1.5.1. má súčet nezávislých náhodných veličín momentovú vytvárajúcu funkciu tvaru $M_S(t) = \prod_{i=1}^{n} M_{X_i}(t)$, kde $M_{X_i}(t)$ je momentová vytvárajúca funkcia náhodnej veličiny $X_i, i = 1, 2, ..., n$. Využijeme, že náhodné veličiny $X_1, ..., X_n$ sú rovnako rozdelené. To znamená, že ich momentové vytvárajúce funkcie sa navzájom rovnajú. Označme

$$M_X(t) = M_{X_1}(t) = \cdots = M_{X_n}(t).$$

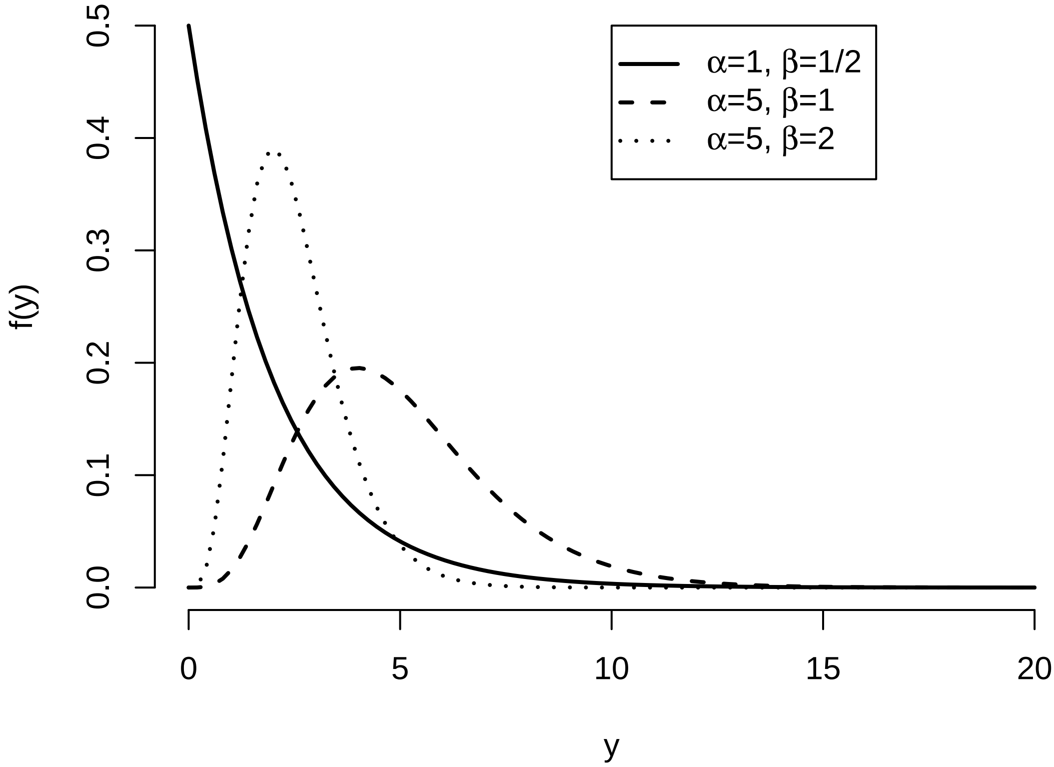
Odtiaľ dostávame

$$M_S(t) = \prod_{i=1}^{n} M_X(t) = (M_X(t))^n.$$

□



Obr. 1.1: Pravdepodobnosť dvoch Poissonových rozdelení s parametrom $\lambda = 4$ a $\lambda = 1$.



Obr. 1.2: Hustoty niekoľkých gama rozdelení.

Kapitola 2

Zložené Poissonovo rozdelenie

V tejto kapitole zavedieme pojem zloženého rozdelenia a popíšeme jeho vlastnosti ako stredná hodnota, rozptyl a momentová vytvárajúca funkcia. V časti 2.2 zadefinujeme zložené Poissonovo rozdelenie a pomocou pripravených vzťahov odvodíme jeho chrakteristiky.

2.1 Zložené rozdelenie

Rozdelenie pravdepodobnosti náhodných veličín, ktoré vzniknú ako súčet náhodného počtu nezávislých rovnako rozdelených náhodných veličín sa nazýva zložené rozdelenie. Zložené rozdelenia sa v praxi uplatňujú pri analýze poisťovacích dát alebo pri modelovaní a predpovedi zrážok. Rada odborníkov sa vo svojich prácach venuje metódam odhadu parametrov zloženého rozdelenia pre konkrétne vybrané rozdelenia náhodných veličín N a $X_1, X_2, ...$ Napríklad v článku Dunn (2004) sa autor zaoberá modelovaním zrážok, kde predpokladá Poissonovo rozdelenie pre počet prehánok za deň (prípadne počet zrážkových dní v týždni, mesiaci a pod.) a že množstvo zrážok má gama rozdelenie.

Definícia 2.1.1. *Zložené rozdelenie je rozdelenie náhodnej veličiny* $S = \sum_{i=1}^{N} X_i$*, kde:*
1. $X_1, X_2, ...$ *je postupnosť vzájomne nezávislých a rovnako rozdelených náhodných veličín,*
2. N *je náhodná veličina nezávislá na* $\{X_i, i = 1, 2, ...\}$*, nadobúdajúca hodnoty* $0, 1, ...$

V nasledujúcej vete uvedieme vzťahy pre výpočet strednej hodnoty a rozptylu zloženého rozdelenia. Tieto vzťahy sa nazývajú Waldove identity alebo Waldove rovnosti.

Veta 2.1.1. *Nech* $\{X_1, ..., X_n, ...\}$ *je postupnosť rovnako rozdelených nezávislých náhodných veličín a* N *je náhodná veličina, ktorá nadobúda nezáporné celočíselné hodnoty. Označme* $\mathsf{E}X = \mathsf{E}X_1 = \cdots$ *Nech* N *a* X_i *sú nezávislé pre* $i = 1, 2, ...$ *a nech existujú konečné momenty* $\mathsf{E}X^2$ *a* $\mathsf{E}N^2$*. Potom stredná hodnota a rozptyl náhodnej veličiny* $S = \sum_{i=1}^{N} X_i$ *je*

$$\mathsf{E}S = \mathsf{E}X\mathsf{E}N, \tag{2.1}$$

$$\mathsf{var}S = \mathsf{E}N\mathsf{var}X + (\mathsf{E}X)^2\mathsf{var}N. \tag{2.2}$$

Dôkaz. Počítajme $\mathsf{E}S$:

$$\mathsf{E}S = \mathsf{E}\left(\sum_{i=1}^{N} X_i\right).$$

Použijeme vzťah (1.4) pre podmienenú strednú hodnotu a vetu o úplnej pravdepodobnosti (viď Zvára a Štěpán (2001) veta 2.1. na strane 28):

$$\mathsf{E}S = \mathsf{E}\left(\mathsf{E}\left(\sum_{i=1}^{N} X_i|N\right)\right) = \sum_n \mathsf{E}\left(\sum_{i=1}^{n} X_i\right)\mathsf{P}\left(N = n\right).$$

Náhodné veličiny $X_1, ..., X_N$ sú nezávislé a rovnako rozdelené, teda platí

$$\mathsf{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathsf{E}X_i = \sum_{i=1}^{n} \mathsf{E}X = n\ \mathsf{E}X.$$

Dosadíme do výpočtu a dostaneme

$$\mathsf{E}S = \sum_n n\ \mathsf{E}X\mathsf{P}\left(N = n\right) = \mathsf{E}X\sum_n n\ \mathsf{P}\left(N = n\right) = \mathsf{E}X\mathsf{E}N.$$

V druhej časti dôkazu použijeme vzťah (1.5) a predpoklad, že náhodné veličiny $X_1, ..., X_N$ sú nezávislé a rovnako rozdelené. Počítajme $\mathsf{var}S$:

$$\begin{aligned}
\mathsf{var}S &= \mathsf{var}\left(\sum_{i=1}^{N} X_i\right)\\
&= \mathsf{E}\left(\mathsf{var}\left(\sum_{i=1}^{N} X_i|N\right)\right) + \mathsf{var}\left(\mathsf{E}\left(\sum_{i=1}^{N} X_i|N\right)\right)\\
&= \mathsf{E}\left(\sum_{i=1}^{N} \mathsf{var}X_i|N\right) + \mathsf{var}\left(\sum_{i=1}^{N} \mathsf{E}X_i|N\right)\\
&= \mathsf{E}\left(N\mathsf{var}X\right) + \mathsf{var}\left(N\mathsf{E}X\right) = \mathsf{E}N\mathsf{var}X + \left(\mathsf{E}X\right)^2\mathsf{var}N.
\end{aligned}$$

□

Stále predpokladáme, že náhodné veličiny $X_1, X_2, \ldots$ sú vzájomne nezávislé a rovnako rozdelené, N je náhodná veličina, ktorá nadobúda nezáporné celočíselné hodnoty a že N a X_i sú nezávislé pre $i = 1, 2, \ldots, N$. Označme $M_X(t) = M_{X_1}(t) = \cdots = M_{X_N}(t)$. Z definície 1.1.2. vieme, že $M_S(t) = \mathsf{E}\left(e^{tS}\right)$. Odvodíme vzťah medzi $M_S(t)$ a momentovými vytvárajúcimi funkciami náhodných veličín X_i :

$$M_S(t) = \mathsf{E}\left(e^{tS}\right) = \mathsf{E}\left(e^{t(X_1+\cdots+X_N)}\right) = \mathsf{E}\left(\prod_{i=1}^{N} \exp\{tX_i\}\right)$$

Analogicky ako v dôkaze vety 2.1.1. použijeme vzťah (1.4)

$$M_S(t) = \mathsf{E}\left(\mathsf{E}\left(\prod_{i=1}^{N} \exp\{tX_i\} \mid N\right)\right) = \mathsf{E}\left(\prod_{i=1}^{N} \mathsf{E}\left(\exp\{tX_i\} \mid N\right)\right)$$

Náhodné veličiny X_i sú nezávislé a rovnako rozdelené. Pomocou vety 1.5.2. dostávame

$$\begin{aligned} M_S(t) &= \mathsf{E}\left(\prod_{i=1}^{N} M_X(t) \mid N\right) = \mathsf{E}\left([M_X(t)]^N\right) = \mathsf{E}\left(\exp\{\ln\,[M_X(t)]^N\}\right) \\ &= \mathsf{E}\left(\exp\{N \ln M_X(t)\}\right) = M_N\left(\ln M_X(t)\right). \end{aligned} \tag{2.3}$$

Momentová vytvárajúca funkcia náhodnej veličiny S je teda vyjadrená ako momentová vytvárajúca funkcia $M_N(r)$ náhodnej veličiny N počítaná v bode $r = \ln M_X(t)$.

Z vety 1.1.1. vieme, že ak existuje k-tý moment $\mathsf{E}S^k$, platí $M_S^{(k)}(t)\big|_{t=0} = \mathsf{E}S^k$. Označme $M_S(t) \equiv M$. Odtiaľ dostávame (viď Mandl a Mazurová (1999) na strane 16):

$$\begin{aligned} \frac{\partial\, log\, M}{\partial t}\Big|_{t=0} &= \frac{M'}{M}\Big|_{t=0} = \mathsf{E}S, \\ \frac{\partial^2\, log\, M}{\partial t^2}\Big|_{t=0} &= \frac{M''M - (M')^2}{M^2}\Big|_{t=0} = \mathsf{E}S^2 - (\mathsf{E}S)^2 = \mathsf{var}S, \\ \frac{\partial^3\, log\, M}{\partial t^3}\Big|_{t=0} &= \frac{(M'''M + M''M' - 2M''M')M^2 - (M''M - (M')^2)2MM'}{M^4}\Big|_{t=0} \\ &= \mathsf{E}S^3 - 3\mathsf{E}S^2\mathsf{E}S + 2(\mathsf{E}S)^3 = \mathsf{E}(S - \mathsf{E}S)^3. \end{aligned}$$

Tieto vzťahy nám dávajú ďalší spôsob pre určenie strednej hodnoty a rozptylu náhodnej veličiny S. Navyše môžeme spočítať koeficient šikmosti γ_1 a prípadne i ďalšie vyššie momenty.

2.2 Zložené Poissonovo rozdelenie a jeho vlastnosti

Predpokladajme, že $\{X_1, ..., X_n, ...\}$ je postupnosť rovnako rozdelených nezávislých náhodných veličín, ktoré majú spojité rozdelenie na intervale $(0, \infty)$, N je náhodná veličina, ktorá nadobúda nezáporné celočíselné hodnoty a že N a X_i sú nezávislé pre $i = 1, 2, ...$

Definícia 2.2.1. *Hovoríme, že náhodná veličina* $S = \sum_{i=1}^{N} X_i$ *má zložené Poissonovo rozdelenie, keď* N *má Poissonovo rozdelenie s parametrom* $\lambda > 0$, *t.j.*

$$\mathsf{P}(N = n) = \frac{\lambda^n}{n!} e^{-\lambda}, n = 0, 1, 2, ...$$

Zložené Poissonovo rozdelenie nepatrí medzi diskrétne ani medzi spojité pravdepodobbnostné rozdelenia. Pre náhodnú veličinu s týmto rozdelením je však charakteristické, že má spojité rozdelenie na kladných číslach a s kladnou pravdepodobnosťou nadobúda nulovú hodnotu. Nulu nadobúda práve, keď $N = 0$. Teda platí

$$\mathsf{P}(S = 0) = \mathsf{P}(N = 0) = e^{-\lambda}.$$

Ako sme uviedli v kapitole 1.4 momentová vytvárajúca funkcia Poissonovho rozdelenia má tvar $M_N(t) = e^{\lambda(e^t-1)}$. Dosadením do (2.3) dostávame momentovú vytvárajúcu funkciu zloženého Poissonovho rozdelenia $M_S(t)$:

$$\begin{aligned} M_S(t) &= M_N\left(\ln M_X(t)\right) = \exp\{\lambda(\exp\{\ln M_X(t)\} - 1)\} = \exp\{\lambda(M_X(t) - 1)\} \\ &= e^{\lambda(M_X(t)-1)}. \end{aligned}$$

Strednú hodnotu a rozptyl zloženého Poissonovho rozdelenia odvodíme pomocou Waldových rovností:

$$\begin{aligned} \mathsf{E}S &= \mathsf{E}X\mathsf{E}N = \lambda\,\mathsf{E}X, \\ \mathsf{var}S &= \mathsf{E}N\mathsf{var}X + (\mathsf{E}X)^2\mathsf{var}N = \lambda\,\mathsf{var}X + \lambda(\mathsf{E}X)^2 = \lambda\left(\mathsf{var}X + \mathsf{E}X)^2\right) \\ &= \lambda\left(\mathsf{E}X^2 - (\mathsf{E}X)^2 + \mathsf{E}X)^2\right) = \lambda\,\mathsf{E}X^2. \end{aligned}$$

Strednú hodnotu a rozptyl môžeme tiež vyjadriť pomocou vzťahov pre k-tú parciálnu deriváciu $\log M_S(t) = \lambda(M_X(t) - 1)$:

$$\begin{aligned} \mathsf{E}S &= \frac{\partial\,\lambda(M_X(t)-1)}{\partial t}\bigg|_{t=0} = \lambda\,M_X'(t)\big|_{t=0} = \lambda\,M_X'(0) = \lambda\,\mathsf{E}X, \\ \mathsf{var}S &= \frac{\partial^2\,\lambda(M_X(t)-1)}{\partial t^2}\bigg|_{t=0} = \frac{\partial\,\lambda M_X'(t)}{\partial t}\bigg|_{t=0} = \lambda\,M_X''(t)\big|_{t=0} \\ &= \lambda\,M_X''(0) = \lambda\,\mathsf{E}X^2. \end{aligned}$$

Vidíme, že v oboch prípadoch sme dostali rovnaké výsledky. Tým istým spôsobom spočítame aj tretí centrálny moment náhodnej veličiny S, čo nám umožní vyjadriť jej koeficient šikmosti:

$$\begin{aligned} \mathsf{E}(S - \mathsf{E}S)^3 &= \frac{\partial^3\,\lambda(M_X(t)-1)}{\partial t^3}\bigg|_{t=0} = \frac{\partial\,\lambda M_X''(t)}{\partial t}\bigg|_{t=0} = \lambda\,M_X''(t)\big|_{t=0} \\ &= \lambda\,M_X'''(t)\big|_{t=0} = \lambda\,M_X'''(0) = \lambda\,\mathsf{E}X^3. \end{aligned}$$

Potom koeficient šikmosti

$$\gamma_1 = \frac{\mathsf{E}(S - \mathsf{E}S)^3}{(\mathsf{var}S)^{3/2}} = \frac{\lambda\,\mathsf{E}X^3}{[\lambda\,\mathsf{E}X^2]^{3/2}} = \frac{\mathsf{E}X^3}{\sqrt{\lambda}\,[\mathsf{E}X^2]^{3/2}}.$$

Uvedieme ešte užitočnú vlastnosť zloženého Poissonovho rozdelenia.

Veta 2.2.1. *Nech $S_1, S_2 \ldots S_n$ sú vzájomne nezávislé náhodné veličiny také, že S_i má zložené Poissonovo rozdelenie s parametrom λ_i a distribučnou funkciou $P_i(x)$, t.j. $S_i = \sum_{k=1}^{N^{(i)}} X_k^{(i)}$, $N^{(i)} \sim Po(\lambda_i)$, $X_k^{(i)}$ má distribučnú funkciu $P_i(x)$. Potom náhodná veličina $S = \sum_{i=1}^n S_i$ má zložené Poissonovo rozdelenie s parametrom*

$$\lambda = \sum_{i=1}^n \lambda_i,$$

a rozdelením daným distribučnou funkciou

$$P(x) = \sum_{i=1}^n \frac{\lambda_i}{\lambda} P_i(x).$$

Dôkaz. Viď Tvrzení II.1. v Mandl a Mazurová (1999). □

Kapitola 3

Využitie v poisťovníctve

Základným pojmom poisťovníctva je riziko, ktoré predstavuje možnosť vzniku poistnej udalosti. Poisťovňa pri vykonávaní svojej činnosti toto riziko na seba preberá, preto musí byť obozretná a zodpovedná v jeho klasifikovaní a hodnotení. Riziko poisťovne vychádza z neistoty, či bude prijaté poistné postačujúce na vyplatenie všetkých poistných plnení. Hodnotenie rizika je oblasť poisťovníctva, v ktorej sa využívajú pokročilé metódy teórie pravdepodobnosti a matematickej štatistiky. Dôležitým faktorom pri riešení rozhodujúcich otázok je znalosť základných charakteristík a rozdelení pravdepodobností celkového poistného plnenia.

Označme S náhodnú veličinu so zloženým rozdelením, ktorá predstavuje celkovú výšku škôd za určité pevne zvolené obdobie (obvykle jeden rok). Cieľom je nájsť a preniesť do praxe jej pravdepodobnostné rozdelenie a základné charakteristiky.

V tejto kapitole sa budeme zaoberať využitím zloženého rozdelenia v oblasti neživotného poistenia, pričom budeme predpokladať konkrétne zvolené rozdelenia pre počet poistných udalostí a výšku škody.

Na modelovanie počtu poistných udalostí použijeme Poissonovo rozdelenie. Ako sme uviedli v kapitole 1, dobre aproximuje výskyt zriedkavých javov pri veľkom počte nezávislých opakovaní experimentu.

Ako najvhodnejšie typy rozdelení výšky poistných plnení sa ukazujú tie, ktoré sú pravostranne zošikmené a konvergencia k nule nie je príliš rýchla. V tejto práci budeme výšku poistných plnení modelovať gama rozdelením. Gama rozdelenie je definované len pre kladné hodnoty. Tento predpoklad je v praxi splnený, pretože poistné plnenia nemôžu nadobúdať záporné hodnoty. Gama rozdelenie je taktiež charakteristické tým, že má ťažký chvost. Práve rozdelenia s ťažkými chvostami sa čoraz viac používajú v poisťovníctve, keďže zahŕňajú aj možnosť extrémnych poistných plnení. Na tento účel sa v praxi využíva aj exponenciálne rozdelenie, na rozdiel od neho však gama rozdelenie závisí na dvoch parametroch, preto je viac flexibilnejšie a pri modelovaní výšky poistných plnení poskytuje širší záber možností.

3.1 Tweedieho zložený Poissonov model

Nech N je náhodná veličina, ktorá reprezentuje počet poistných udalostí v sledovanom období (zvyčajne jeden rok) a X_i predstavuje výšku škody $i-$tej poistnej udalosti. Predpokladáme, že výška každej škody má gama rozdelenie s parametrami α a β a že N má Poissonovo rozdelenie s parametrom λ. Celková výška škôd S vznikne ako „Poissonov“ súčet gama-rozdelených náhodných veličín, teda

$$S = X_1 + X_2 + \cdots + X_N = \sum_{i=1}^{N} X_i.$$

Z týchto predpokladov vyplýva, že pravdepodobnosť, že v danom období nenastane žiadna poistná udalosť a celková výška škôd bude nulová, je kladná (v prípade, že $N = 0$):

$$\mathsf{P}(S = 0) = \mathsf{P}(N = 0) = e^{-\lambda}. \tag{3.1}$$

Poznámka 3.1.1. V praxi nemusí byť N vždy známe. Často sa stáva, že poisťovňa má k dispozícii iba celkovú výšku škôd S.

Výsledné rozdelenie nemá jednotné pomenovanie. Ako uvádza Dunn (2004), nazýva sa zložené Poissonovo rozdelenie, zložené gama rozdelenie alebo tiež Poissonovo-gama rozdelenie. Podľa Wüthrich (2003) bol pravdepodobne prvý, ktorý v roku 1984 študoval zložený Poissonov model s použitím gama rozdelenia, štatistik Liverpoolskej univerzity, Maurice Charles Kenneth Tweedie. Preto je model často v literatúre uvádzaný pod názvom Tweedieho zložený Poissonov model, viď napr. Smyth a Jørgensen (2002). Ako naznačuje názov kapitoly, v tejto práci budeme používať posledné zo spomenutých pomenovaní. Skrátene ho budeme nazývať Tweedieho rozdelenie, aj keď v niektorej literatúre pod týmto pojmom rozumieme niečo všeobecnejšie.

Strednú hodnotu a rozptyl Tweedieho rozdelenia získame dosadením do Waldových rovností (2.1) a (2.2). Dostaneme

$$\mathsf{E}S = \lambda \frac{\alpha}{\beta}, \tag{3.2}$$

$$\mathsf{var}S = \lambda \frac{\alpha}{\beta^2} + \lambda \frac{\alpha^2}{\beta^2} = \frac{\lambda\,\alpha}{\beta^2}(1 + \alpha). \tag{3.3}$$

Pri tvorbe rezerv je pre poisťovňu dôležité disponovať určitou informáciou o očakávanej výške škôd. Táto informácia sa získava pozorovaním dát z minulých období, z ktorých sa vhodnou matematickou metódou stanoví očakávaná výška škôd. V jazyku teórie pravdepodobnosti a štatistiky to znamená určiť strednú hodnotu náhodnej veličiny S. Ak by sme poznali skutočné hodnoty parametrov α, β a λ, nemali by sme problém ju vypočítať pomocou Waldových rovností. Poisťovňa však tieto hodnoty k dispozícii nemá. V obdobných prípadoch sa v praxi aplikujú postupy matematickej štatistiky a metódy odhadov parametrov. Chceme teda odhadovať strednú hodnotu, ktorá podľa (3.2) závisí na všetkých troch parametroch Tweedieho rozdelenia. To by v praxi znamenalo, že budeme odhadovať parametre α, β a λ. Nás však zaujíma až výsledná stredná hodnota. Preto si označíme

$$\mu = \mathsf{E}S = \lambda \frac{\alpha}{\beta} \tag{3.4}$$

a budeme priamo odhadovovať parameter μ.

Postup odahadu parametra μ popíšeme v časti 3.2 Odhady parametrov. Najskôr zavedieme alternatívnu parametrizáciu Tweedieho rozdelenia, ktorá je určená taktiež trojicou parametrov, ktoré označíme μ, ϕ a p. Paramater μ sme zaviedli rovnosťou (3.4). Ako uvádza Smyth a Jørgensen (2002) (viď tiež (3.10)) takto skonštruované Tweedieho rozdelenie patrí do exponenciálnej rodiny rozdelení a rozptyl S sa dá napísať ako

$$\mathsf{var} S = \mu^p \, \phi, \tag{3.5}$$

kde ϕ je tzv. disperzný parameter a

$$p = \frac{\alpha + 2}{\alpha + 1}. \tag{3.6}$$

Ako vieme parameter α je kladný, z toho vyplýva, že $1 < p < 2$.

Tým sme dostali novú parametrizáciu Tweedieho rozdelenia s parametrami μ, ϕ a p, so strednou hodnotou μ a rozptylom $\mu^p \, \phi$. Výhodou tejto parametrizácie je, že sa sústredí na pre prax dôležitý parameter, ktorým je stredná hodnota μ.

Keď dáme do rovnosti vzťahy (3.3) a (3.5) pre rozptyl, môžeme vyjadriť ϕ pomocou trojice parametrov (λ, α, β) :

$$\phi = \frac{\mathsf{var} S}{\mu^p} = \frac{\lambda \, \alpha \, (1 + \alpha)}{\beta^2 \, \mu^p}.$$

Dosadíme $\mu = \lambda \dfrac{\alpha}{\beta}$ a dostávame:

$$\phi = \frac{\lambda \, \alpha \, (1 + \alpha)}{\beta^2} \frac{(\lambda \, \alpha)^{-p}}{\beta^{-p}} = \frac{1}{(\lambda \, \alpha)^{p-1} \, \beta^{2-p} \, (p-1)}.$$

Môžeme vidieť, že ak sa zväčší intenzita vzniku poistných udalostí λ, bez toho, aby sa zmenila priemerná výška škôd, hodnota disperzného parametra ϕ sa zníži, zatiaľ čo stredná hodnota μ sa zvýši.

Teraz odvodíme vzťahy pre λ a β v závislosti na parametroch μ, p, ϕ. Vieme, že $\mu = \lambda \dfrac{\alpha}{\beta}$, z toho $\beta = \dfrac{\lambda \, \alpha}{\mu}$ a $\lambda \, \alpha = \mu \, \beta$. Dosadíme do $\phi = \dfrac{\lambda \, \alpha \, (1 + \alpha)}{\beta^2 \, \mu^p}$ a počítame:

$$\phi = \lambda \, \alpha \, (\alpha + 1) \, \frac{\mu^2}{(\lambda \, \alpha)^2 \, \mu^p} = \frac{\alpha + 1}{\lambda \, \alpha \, \mu^{p-2}},$$

odtiaľ dostávame

$$\lambda = \frac{\alpha + 1}{\alpha \, \mu^{p-2} \, \phi} = \frac{\mu^{2-p}}{\phi \, (2-p)}. \tag{3.7}$$

Parameter β vyjadríme analogicky:

$$\phi = \frac{\mu \, \beta \, (1 + \alpha)}{\beta^2 \, \mu^p} = \frac{(\alpha + 1)}{\beta \, \mu^{1-p}},$$

a teda

$$\beta = \frac{\alpha + 1}{\phi \, \mu^{p-1}} = \frac{1}{\phi \, (p-1) \, \mu^{p-1}}. \tag{3.8}$$

Parameter α vyjadríme jednoduchými úpravami vzťahu (3.6) a dostaneme

$$\alpha = \frac{2-p}{p-1}. \tag{3.9}$$

3.2 Odvodenie hustoty

Predpokladajme, že nastalo $n > 0$ poistných udalostí, ktorých výšky škôd sú nezávislé. Z vlastností gama rozdelenia teda vyplýva, že podmienené rozdelenie celkovej výšky škôd S, keď je daný počet poistných udalostí $N = n$, má tiež gama rozdelenie s parametrami $n\alpha$ a β, t.j.

$$f(s \mid n) = \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} s^{n\alpha-1} e^{-\beta s} \mathbb{I}_{(0,\infty)}(s).$$

Z vlastností združenej a podmienej hustoty (viď Anděl (2007) na strane 54) vieme, že

$$f(s \mid n) = \frac{f_{S,N}(s,n)}{f_N(n)},$$

kde $f_{S,N}(s,n)$ je združená hustota náhodných veličín S a N a $f_N(n)$ je rozdelenie pravdepodobnosti náhodnej veličiny N. Z toho vyplýva, že $f(s,n) = f(s \mid n)\, f(n)$. Dosadením do týchto vzťahov dostávame

$$f_{S,N}(s,n) = \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} s^{n\alpha-1} e^{-\beta s} \frac{\lambda^n}{n!} e^{-\lambda} = a\,(n,s,\phi,p)\, \exp\left\{\frac{1}{\phi} t\,(s,\mu,p)\right\},$$

kde

$$\begin{aligned} a\,(n,s,\phi,p) &= \frac{1}{n!\Gamma(n\alpha)s} \left(s^{\alpha} \lambda\, \beta^{\alpha}\right)^n \\ &= \frac{1}{n!\Gamma(n\alpha)s} \left[\frac{\mu^{2-p}}{\phi(2-p)} \frac{1}{\phi^{\alpha}(p-1)^{\alpha}(\mu^{p-1})^{\alpha}} s^{\alpha}\right]^n = \\ &= \frac{1}{n!\Gamma(n\alpha)s} \left[\frac{(1/\phi)^{\alpha+1} s^{\alpha}}{(p-1)^{\alpha}(2-p)}\right]^n, \end{aligned}$$

pretože $\dfrac{\mu^{2-p}}{(\mu^{p-1})^{\alpha}} = \dfrac{\mu^{2-p}}{(\mu^{p-1})^{\frac{p-2}{1-p}}} = \dfrac{\mu^{2-p}}{\mu^{2-p}} = 1.$

Ďalej

$$\begin{aligned} \exp\left\{-\beta s - \lambda\right\} &= \exp\left\{-s\frac{1}{\phi(p-1)\mu^{p-1}} - \frac{\mu^{2-p}}{\phi(2-p)}\right\} = \\ &= \exp\left\{\frac{1}{\phi}\left[s\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right]\right\}, \end{aligned}$$

teda $t\,(s,\mu,p) = s\dfrac{\mu^{1-p}}{1-p} - \dfrac{\mu^{2-p}}{2-p}$.

Združená hustota náhodných veličín S a N má nasledujúci tvar:

$$f_{S,N}(s,n) = \begin{cases} \exp\left\{-\frac{\mu^{2-p}}{\phi\,(2-p)}\right\}, & \text{pre } n = 0 \text{ a } s \in (0,\infty), \\ a\,(n,s,\phi,p)\, \exp\left\{\frac{1}{\phi} t\,(s,\mu,p)\right\}, & \text{pre } n > 0 \text{ a } s \in (0,\infty)\ , \\ 0, & \text{inak.} \end{cases}$$

Hustotu S na $(0,\infty)$ potom dostaneme sčítaním cez všetky možné hodnoty N :

$$\begin{aligned}
f_S(s) &= \sum_n f_{S,N}(s,n), \\
&= \sum_n a\,(n,s,\phi,p)\, \exp\left\{\frac{1}{\phi} t\,(s,\mu,p)\right\}, \\
&= c\,(s,\phi,p)\, \exp\left\{\frac{1}{\phi} t\,(s,\mu,p)\right\}, \\
&= c\,(s,\phi,p)\, \exp\left\{\frac{1}{\phi}\left[s\theta(\mu) - \kappa(\theta(\mu))\right]\right\}, \qquad (3.10)
\end{aligned}$$

kde

$$\begin{aligned}
c\,(s,\phi,p) &= \sum_n \frac{1}{n!\Gamma(n\alpha)s}\left[\frac{(1/\phi)^{\alpha+1}s^{\alpha}}{(p-1)^{\alpha}(2-p)},\right]^n, \\
\theta(\mu) &= \frac{\mu^{1-p}}{1-p} \qquad \text{a} \qquad \kappa(\theta(\mu)) = \frac{\mu^{2-p}}{2-p}.
\end{aligned}$$

Zápis hustoty (3.10) dokazuje, že toto rozdelenie patrí do exponenciálnej rodiny rozdelení.

Náhodná veličina S teda nadobúda nulovú hodnotu s pravdepodobnosťou

$$\mathsf{P}(S=0) = \exp\left\{-\frac{\mu^{2-p}}{\phi\,(2-p)}\right\},$$

a na intervale $(0,\infty)$ má hustotu

$$f_S(s) = c\,(s,\phi,p)\, \exp\left\{\frac{1}{\phi} t\,(s,\mu,p)\right\}.$$

Poznámka 3.2.1. Hustota $f_S(s)$ by sa dala vyjadriť tiež pomocou parametrov α, β a λ použitím vzťahov (3.7),(3.8) a (3.9).

Príklad hustoty Tweedieho rozdelenia s konkrétne zvolenými parametrami je uvedený na Obr. 3.1.

3.3 Odhady parametrov

V tomto odstavci uvedieme bodové odhady strednej hodnoty μ a disperzného parametra ϕ, pričom pri každom použijeme inú metódu, a to metódu maximálnej vierohodnosti pre μ a momentovú metódu pre ϕ. Najskôr uvedieme a vysvetlíme niekoľko označení, ktoré budeme používať.

Nech $S_1, S_2, \ldots, S_m$ je náhodný výber z Tweedieho rozdelenia s parametrami μ, ϕ a p. Funkcia náhodného výberu

$$\bar{S}_m = \frac{1}{m}\sum_{i=1}^{m} S_i$$

sa nazýva výberový priemer a funkcia

$$S_m^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(S_i - \bar{S}_m\right)^2$$

sa nazýva výberový rozptyl.
Výberový priemer je nestranný a konzistentný (viď Definice 5.4. a 5.5. v Dupač a Hušková (2001)) odhad strednej hodnoty rovnako ako výberový rozptyl je nestranný a konzistentný odhad rozptylu (viď Věta 5.1. v Dupač a Hušková (2001)).

Ďalej budeme predpokladať, že poznáme hodnotu parametra p alebo jeho odhad, ktorého postup stručne popíšeme v časti 3.3.3.

3.3.1 Odhad parametra μ

Metóda maximálnej vierohodnosti je jednou z najpoužívanejších metód určovania bodových odhadov. Cieľom tejto metódy je odhadnúť parametre $\boldsymbol{\theta}$ tak, aby bola pri každej hodnote náhodného výberu $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ maximalizovaná funkcia vierohodnosti

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}).$$

Funkcia vierohodnosti sa rovná združenej hustote $f(\boldsymbol{x}; \boldsymbol{\theta})$ náhodných veličín X_1, $X_2, \ldots, X_n$, vyšetrujeme ju ale ako funkciu parametrov $\boldsymbol{\theta}$. Pri hľadaní maximálne vierohodných odhadov je výhodnejšie pracovať s logaritmom funkcie vierohodnosti, teda s logaritmickou funkciou vierohodnosti

$$\ell(\boldsymbol{\theta}) = \log\left\{L(\boldsymbol{\theta})\right\} = \sum_{i=1}^{n} \log\left\{f(x_i; \boldsymbol{\theta})\right\},$$

ktorá nadobúda svoje maximum v rovnakom bode ako $L(\boldsymbol{\theta})$. Označme $\widehat{\boldsymbol{\theta}}$ hodnotu, ktorá maximalizuje $\ell(\boldsymbol{\theta})$. Ak je táto funkcia dostatočne hladká, potom je možné hľadanie maximálnych hodnôt zjednodušiť na prosté hľadanie maxím pomocou parciálnych derivácií, ktoré vedie k riešeniu rovnice

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta}) = 0.$$

Majme náhodný výber $S_1, S_2, \ldots, S_m$ z Tweedieho rozdelenia s parametrami μ, ϕ a p. Metódou maximálnej vierhodnosti budeme odhadovať parameter μ, teda strednú hodnotu. Najskôr zostavíme funkciu vierohodnosti:

$$L(\mu, \phi, p) = \prod_{i=1}^{m} f(s_i, \mu, \phi, p) =$$
$$= \prod_{i=1}^{m} c\left(s_i, \phi, p\right) \exp\left\{\frac{1}{\phi} \sum_{i=1}^{m} \left(s_i \frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right)\right\},$$

potom logaritmická funkcia vierohodnosti $\ell(\mu, \phi, p)$ má tvar

$$\ell(\mu, \phi, p) = \sum_{i=1}^{m} \log c\,(s_i, \phi, p) + \frac{1}{\phi} \sum_{i=1}^{m} \left(s_i \frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p} \right).$$

Deriváciu logaritmickej funkcie vierohodnosti položíme rovnú nule a dostaneme tak rovnicu pre hľadaný parameter:

$$\frac{\partial \ell(\mu, \phi, p)}{\partial \mu} = \frac{1}{\phi} \frac{\partial}{\partial \mu} \sum_{i=1}^{m} t(s_i, \mu, p) = \frac{1}{\phi} \sum_{i=1}^{m} \left(\frac{s_i - \mu}{\mu^p} \right)$$

$$\begin{aligned}
\frac{1}{\phi} \sum_{i=1}^{m} \left(\frac{s_i - \mu}{\mu^p} \right) &= 0 \\
\frac{1}{\mu^p} \sum_{i=1}^{m} (s_i - \mu) &= 0 \\
\sum_{i=1}^{m} s_i - m\mu &= 0 \\
\widehat{\mu} = \frac{1}{m} \sum_{i=1}^{m} s_i &= \bar{S}_m, \qquad (3.11)
\end{aligned}$$

kde $\bar{S}_m$ je výberový priemer. Touto metódou sme dospeli k maximálne vierohodnému odhadu parametra μ, ktorý je nestranným a konzistentným odhadom strednej hodnoty.

3.3.2 Odhad parametra ϕ

Opäť predpokladáme náhodný výber $S_1, S_2, \ldots, S_m$ z Tweedieho rozdelenia s parametrami μ, ϕ a p. Momentová metóda využíva, že máme k dispozícii konzistentné odhady momentov, ktoré zvyčajne vieme vyjadriť ako funkcie neznámych parametrov. To znamená, že vyjadríme strednú hodnotu $\mathsf{E}S_i = g_1(\mu, \phi, p)$ a rozptyl $\mathsf{var}S_i = g_2(\mu, \phi, p)$. Vieme, že $\bar{S}_m$ a S_m^2 sú konzistentné odhady $\mathsf{E}S_i$ a $\mathsf{var}S_i$. Budeme riešiť sústavu rovníc

$$\begin{aligned}
g_1(\widehat{\mu}, \widehat{\phi}, p) &= \bar{S}_m, \\
g_2(\widehat{\mu}, \widehat{\phi}, p) &= S_m^2.
\end{aligned}$$

V používanom modeli platí $\mathsf{E}S_i = \mu$ a $\mathsf{var}S_i = \mu^p \phi$, dosadíme do sústavy a dostaneme

$$\begin{aligned}
\widehat{\mu} &= \bar{S}_m, \\
\widehat{\mu}^p \widehat{\phi} &= S_m^2.
\end{aligned}$$

Z rovníc vyjadríme $\widehat{\phi}$ a získame odhad disperzného parametra ϕ

$$\widehat{\phi} = \frac{S_m^2}{\bar{S}_m^p} = \frac{S_m^2}{\widehat{\mu}^p}. \qquad (3.12)$$

Odhad $\widehat{\phi}$ môžeme vyjadriť ako funkciu h premenných $\bar{S}_m$ a S_m^2, kde

$$h(\bar{S}_m, S_m^2) = \frac{S_m^2}{\bar{S}_m^p}.$$

Vidíme, že funkcia h je spojitá. Keďže výberový priemer aj výberový rozptyl sú konzistentné odhady, z vety o spojitej transormácii výplýva, že $\widehat{\phi}$ je konzistentný odhad disperzného parametra.

3.3.3 Odhad parametra p

Rovnako ako v prvých dvoch prípadoch predpokladáme náhodný výber S_1, $S_2, \ldots, S_m$ z Tweedieho rozdelenia s parametrami μ, ϕ a p a naším cieľom je nájsť odhad týchto parametrov. Ak by sme postupovali metódou maximálnej vierohodnosti, dostali by sme odhad parametrov ako

$$(\widehat{\mu}, \widehat{\phi}, \widehat{p}) = \mathsf{arg\ max}_{\mu,\phi,p}\, \ell(\mu, \phi, p) = \mathsf{arg\ max}_{\mu,\phi,p} \sum_{i=1}^{m} \log f(s_i, \mu, \phi, p).$$

To však vedie k výrazu, ktorý je zložité maximalizovať. Z tohto dôvodu sa parameter p odhaduje mierne modifikovaným postupom maximálnej vierohodnosti, tzv. metódou profilovej vierohodnosti (anglicky *profile likelihood*), ktorú je v niektorých prípadoch jednoduchšie vyhodnotiť.

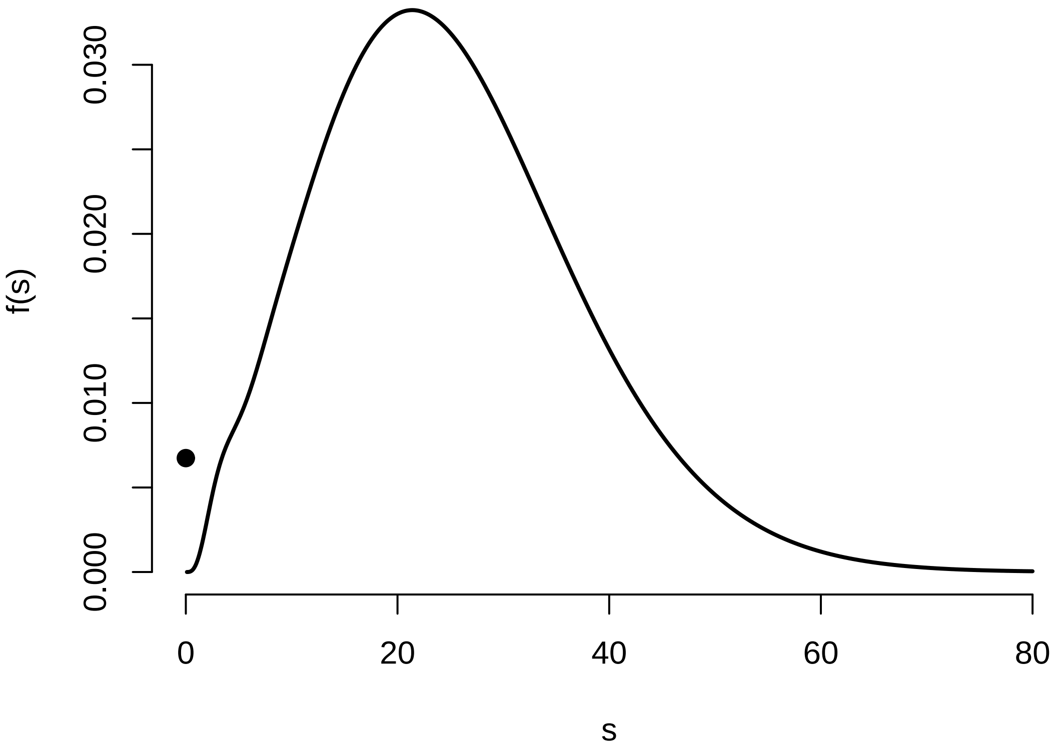
Predpokladajme, že parameter p je daný. Z (3.11) vidíme, že odhad $\widehat{\mu}$ nie je závislý na p, zatiaľ čo $\widehat{\phi}$ na p závisí (viď (3.12)). Označme $\widehat{\phi}_p$ odhad disperzného parametra pre dané p. Pre každé p spočítame $\ell(\widehat{\mu}, \widehat{\phi}_p, p)$ a jeho odhad nájdeme ako maximum tejto funkcie vzhľadom k p, t.j.

$$\widehat{p} = \mathsf{arg\ max}_p \ell(\widehat{\mu}, \widehat{\phi}_p, p).$$

V praxi sa prechádza sieť možných hodnôt p z intervalu $(1, 2)$ a po vyčíslení jeho odhadu, sa hodnoty $\widehat{\mu}$ a $\widehat{\phi}$ vyhodnocujú znovu podľa postupov popísaných v častich 3.3.1 a 3.3.2.

V používanom modeli platí, že $p \in (1, 2)$. Pri jeho odhadovaní by sa teoreticky mohlo stať, že by odhad ležal mimo tohto intervalu. To by znamenalo, že by sme nemohli použiť Tweedieho rozdelenie v otázke tvroby rezerv (napr. ak $p = 0$, ide o normálne rozdelenie). Wüthrich (2003) však uvádza, že vo väčšine aplikácií z poisťovníctva, odhad parametra p ležal vždy presne v intervale $(1, 2)$, čo dokazuje aj jednoduchá analýza dát uvedená v kapitole 4.

Bližší popis metódy profilovej vierohodnosti je možné nájsť v Young a Smith (2005) v časti 8.6.



Obr. 3.1: Hustota Tweedieho rozdelenia s parametrami $\lambda = 5, \alpha = 5, \beta = 1$.

Kapitola 4

Analýza dát

V predchádzajúcich kapitolách sme sa venovali prevažne teoretickému spracovaniu zloženého Poissonovho rozdelenia. Z uvedeného dôvodu sa v tejto kapitole budeme zaoberať analýzou konkrétnych dát, na ktorých sa snažíme ilustrovať popísaný model v praxi. Použité dáta sú sprístupnené na internetovej adrese http://www.statsci.org/data/general/motorins.html a vzťahujú sa k poisteniu motorových vozidiel vo Švédsku.

Dáta popisujú viacero aspektov, ktoré sú potrebné k čo najpodrobnejšej analýze. Pozorujeme 7 náhodných veličín:

- *Kilometres* - počet najazdených kilometrov za rok.
- *Zone* - geografická oblasť Švédska.
- *Bonus* - číslo 1 až 7 označuje kategóriu, do ktorej sú poistení zaradení podľa počtu rokov od poslednej poistnej udalosti.
- *Make* - číslo 1 až 9 označuje model vozidla. Modely nie sú bližšie špecifikované, aby analýza neovplyvnila ich predaj.
- *Insured* - počet poistených v danom období.
- *Claims* - počet poistných udalostí.
- *Payment* - celková výška škody v švédskych korunách.

Smyth a Jørgensen (2002) uvažujú a analyzujú dáta pre Zónu 1. My sa zameriame na Zónu 5, ktorá pozostáva z malých miest a okolitých častí severného Švédska a na Zónu 6, ktorá zahŕňa vidiecke oblasti na severe Švédska. Z oboch zón vylúčime z kategórie, ktorá popisuje model vozidla, triedu 9, keďže ju tvoria rozličné, nešpecifikované typy vozidiel, ktoré nie sú súčasťou predchádzajúcich ôsmych tried. Vo vyššie uvedenom článku autori uvažujú taktiež vplyv vysvetľujúcich premenných. V tejto práci uplatníme zjednodušný prístup k analýze a vplyv vystvetľujúcich premenných zanedbáme.

V Zóne 5 bolo zaznamenaných 1 549 poistných udalostí v 278 kategóriách a z týchto kategórií bolo 77 bez poistných udalostí. Keď dáme do pomeru počet kategórií bez poistnej udalosti a celkový počet kategórií, môžeme vyčísliť pravdepodobnosť, že nenastala žiadna poistná udalosť, ktorá vychádza približne 0,277. Túto hodnotu budeme porovnávať s hodnotou, ktorú získame z (3.1) po odhadnutí parametra λ. Priemerná výška škody je 29 281 švédskych korún (ďalej len

„SEK") a najvyššia škoda, ktorá bola vyplatená pri jednej poistnej udalosti činí 604 369 SEK.

V Zóne 6 nastalo 2 983 poistných udalostí v 280 pozorovaných kategóriách, z ktorých nenastala žiadna poistná udalosť v 51 kategóriách. Pravdepodobnosť, že poistná udalosť nenastane je 0,182. Priemerná výška škody je 58 813 SEK, najvyššia 1 148 035 SEK. Histogramy výšky škôd pre jednotlivé zóny sú zobrazené na Obr. 4.1.

Pre lepšiu prehľadnosť získané hodnoty uvádzame v nasledujúucej tabuľke:

	Zóna 5	**Zóna 6**
Kategórie	278	280
Poistné udalosti	1 549	2 983
Kategórie bez p. udalostí	77	51
Priemerná škoda v SEK	29 281	58 813
Najvyššia škoda v SEK	604 369	1 148 035

Tabuľka 4.1: Základné údaje Zóny 5 a Zóny 6.

Naším cieľom je overiť, či je možné na celkovú výšku škôd uplatniť Tweedieho zložený Poissonov model a využiť jeho základné charakteristiky k popísaniu predpokladaného škodného priebehu v ďalších rokoch. Z tohto dôvodu bude naším bodom záujmu náhodná veličina *Payment,* ktorá modeluje práve celkovú výšku škôd v SEK. Podľa značenia použitého v kapitole 3 predstavuje *Payment* náhodnú veličinu S a *Claims* náhodnú veličinu N. Pre jednoduchší výpočet sme hodnoty *Payment* vydelili tisíckou, t.j. hodnoty sú v 1 000 SEK. Odhady strednej hodnoty μ, disperzného prametra ϕ a paramtera p získame z dát postupom, ktorý sme popísali v časti 3.3 Odhady parametrov. Použitím vzťahov (3.7),(3.8) a (3.9) potom dostaneme tiež hodnoty parametrov λ, α a β.

Odhady parametrov sú uvedené v nasledujúcej tabuľke:

	Zóna 5	**Zóna 6**
μ	29,28094	58,81316
ϕ	12,88332	13,44896
p	1,68776	1,72857
λ	0,71353	0,82783
α	0,45401	0,37255
β	0,01106	0,00524

Tabuľka 4.2: Parametre záskané z dát.

Parametre α a β sú parametrami gama rozdelenia náhodnej veličiny X_i, ktorá predstavuje výšku škody pre i-tú poistnú udalosť. Môžeme teda skonštruovať hustotu X_i pre jednotlivé zóny (viď Obr. 4.2).

Jedným z predpokladov Tweedieho rozdelenia je, že parameter p musí ležať v intervale $(1, 2)$. Tento predpoklad je splnený (viď Tabuľka 4.2), môžeme teda tento model použiť. Pomocou odhadnutých parametrov zostrojíme príslušné hustoty pre Zónu 5 a Zónu 6 a preložíme nimi histogram výšky škôd, aby sme mohli porovnať skutočnosť s použitým modelom. Hustoty jednotlivých rozdelení spolu s histogramom sú zobrazené na Obr. 4.3. Pravdepodobnosť, že poistná udalosť nenastane, teda že celková výška škôd bude nulová (viď Tabuľka 4.3), sa zdá byť modelom nadhodnotená, preto nie je v obrázku uvedená.

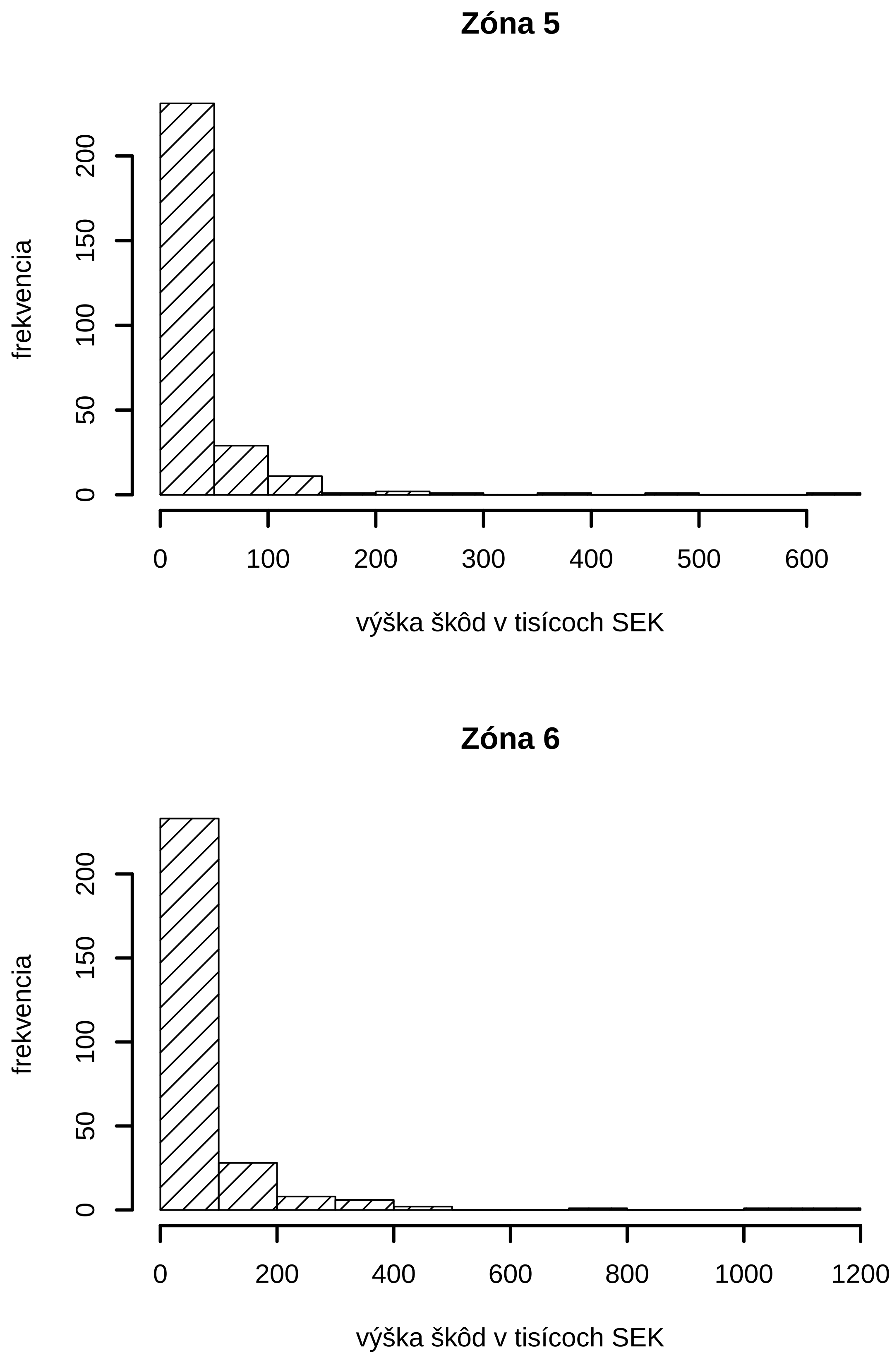
Ako sme uviedli vyššie, budeme ešte porovnávať dva postupy výpočtu pravdepodobnosti, že nenastane žiadna poistná udalosť. Tieto hodnoty sú uvedené v Tabuľke 4.3, kde hodnota Pravdepodobnosť 1 je vypočítaná ako pomer počtu kategórií bez poistnej udalosti a celkového počtu kategórií a Pravdepodobnosť 2 je vyčíslená použitím vzťahu (3.1), do ktorého sme dosadili odhad parametra λ. Ako vidíme tieto hodnoty sa výrazne líšia. Pravdepodobnosť 2, ktorú sme získali použitím Tweedieho zloženého Poissonvho modelu je nadhodnotená. Tento rozdiel môže byť spôsobený tým, že sme analýzu dát zjednodušili a neuvažovali sme vplyv vysvetľujúcich premenných.

	Zóna 5	**Zóna 6**
Pravdepodobnosť 1	0,277	0,1821
Pravdepodobnosť 2	0,4899	0,437

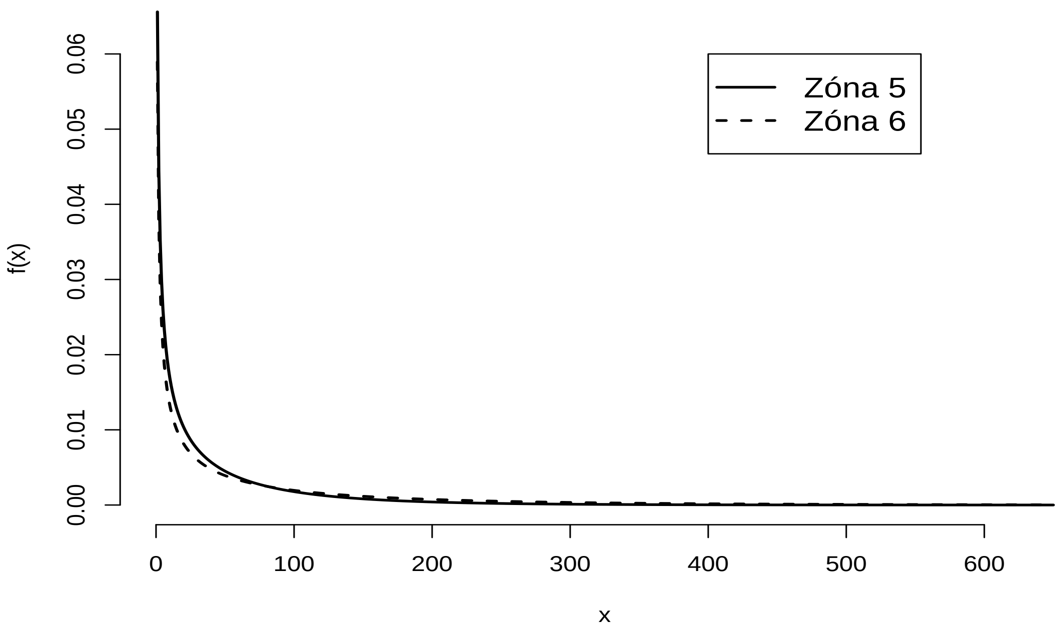
Tabuľka 4.3: Porovnanie pravdepodobností, že nenastane poistná udalosť.

Cieľom tejto kapitoly bolo ilustrovať Tweedieho zložený Poissonov model na reálnych dátach. Pri analýze sme sa zamerali na Zónu 5 a Zónu 6 a zvolili sme zjednodušený postup, pri ktorom sme zanedbali vplyv vysvetľujúcich premenných. Z tohto dôvodu sa skutočnosť nie úplne zhoduje s použitým modelom. Pri analýze sme použili postupy popísané v teoretickej časti práce, uviedli sme hodnoty parametrov pre oba typy parametrizácie tohto rozdelenia a pre dve rôzne metódy výpočtu sme porovnali pravdepodobnosť, že nenastane žiadna poistná udalosť.

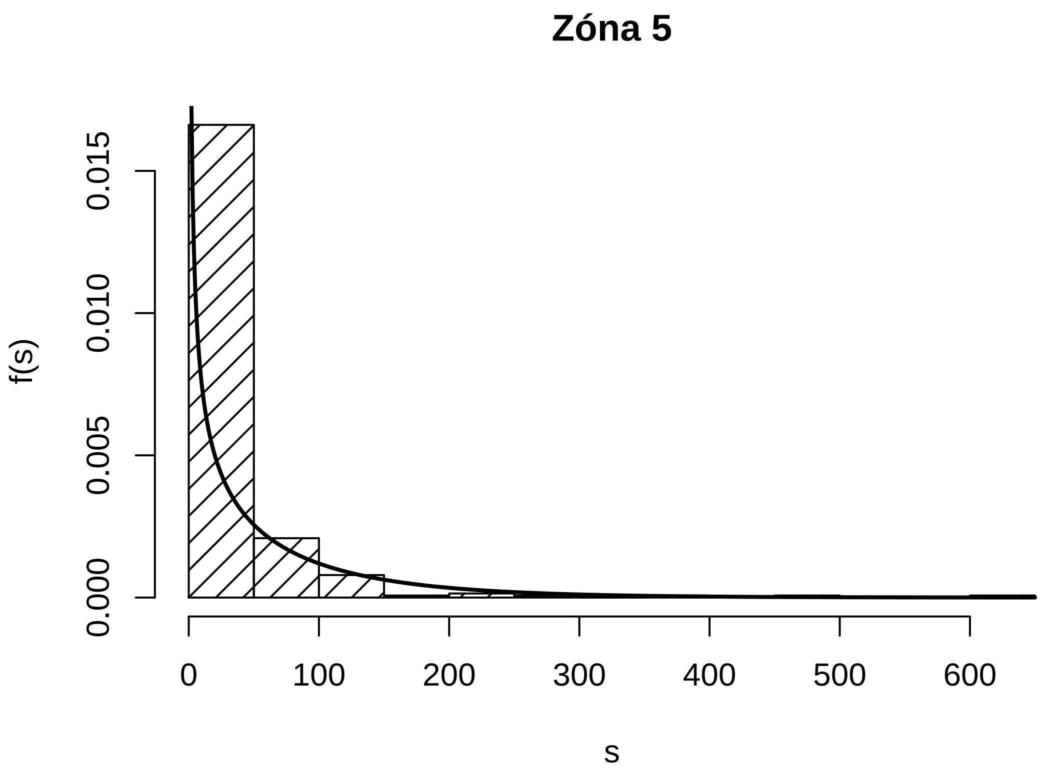
Pri dôkladnejšom pozorovaní by sme museli určiť vysvetľujúce premenné a zahrnúť ich vplyv do analýzy. Okrem vyššie uvedených výsledkov by sme mohli taktiež stanoviť pravdepodobnosť, že celková výška škody presiahne v danom období určitú hranicu.

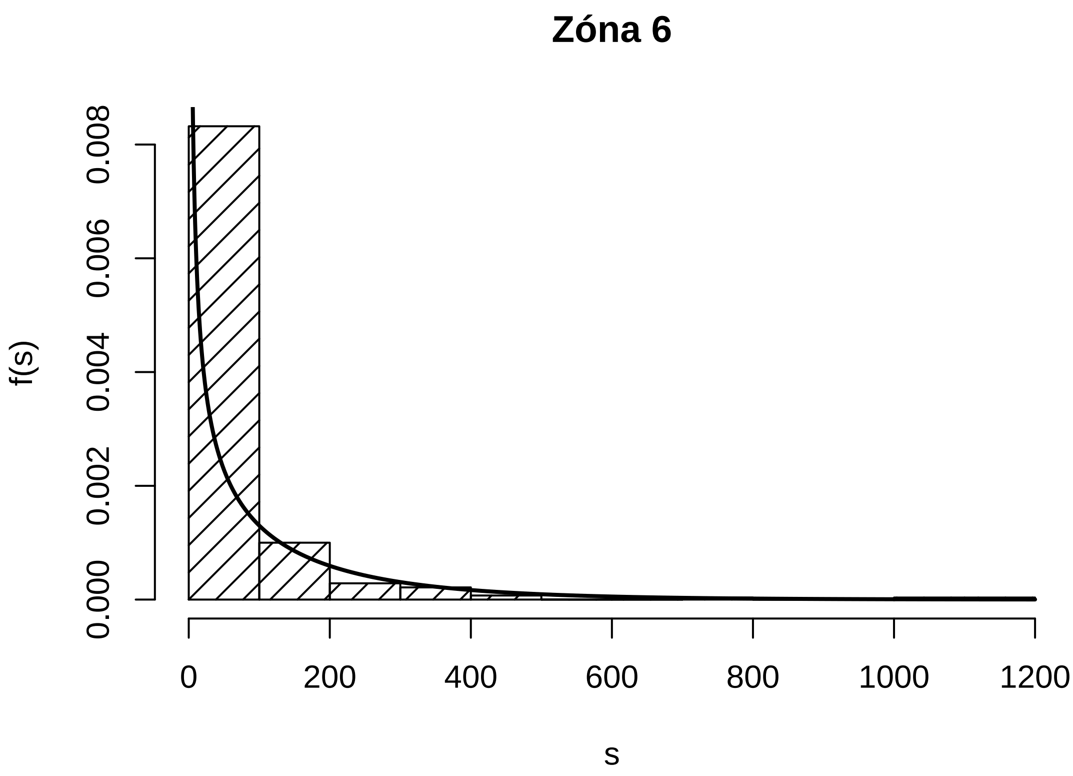


Obr. 4.1: Histogramy výšky škôd pre jednotlivé zóny.



Obr. 4.2: Hustoty náhodnej veličiny X_i pre jednotlivé zóny.





Obr. 4.3: Histogramy preložené hustotou.

Záver

V bakalárskej práci sme sa zaoberali zloženým Poissonovým rozdelením, jeho vlastnosťami a možným využitím, pričom sme sa zamerali najmä na aplikáciu v poisťovníctve. Spracovaním jednotlivých kapitol sme sa snažili splniť stanovený cieľ.

V prvej kapitole sme rozpracovali základné pojmy, vety a tvrdenia, ktoré boli východiskom pre vysvetlenie ďalších postupov.

Druhá kapitola bola venovaná zloženému rozdeleniu, ktoré sme najskôr definovali na všeobecnej úrovni a následne sme popísali vlastnosti zloženého Poissonovho rozdelenia.

Využitím v praxi sme sa zaoberali v tretej kapitole. Uvažovali sme konrétne rozdelenia pre počet poistných udalostí a jednotlivé výšky škôd, popísali sme postup odvodenia hustoty takto skonštruovaného rozdelenia a odhad parametrov.

Posledná kapitola bola venovaná analýze dát, kde sme na reálnych dátach ilustrovali aplikáciu Tweedieho zloženého Poissonovho modelu.

Spracovaním bakalárskej práce sme nevyčerpali všetky aspekty danej témy. Použitie Tweedieho zloženého Poissonovho modelu je podrobnejšie popísané vo vyššie spomenutých publikáciách Wüthrich (2003) a Smyth a Jørgensen (2002). Autori v nich uvažujú vplyv vysvetľujúcich premenných a pri analýze dát aplikujú pokročilejšie štatistické metódy ako sú napr. zovšeobecnené lineárne modely. Metódy stanovenia rezerv, ktorých základom je využitie zloženého rozdelenia, sú uvedené taktiež v Wüthrich a Merz (2008), kde je uplatnenie Tweedieho zloženého Poissonovho modelu zhrnuté v časti 5.2.5. Spektrum tejto témy ponúka priestor na hlbšie štúdium jej využitia a porovnania s inými používanými metódami v oblasti poisťovníctva.

Zoznam použitej literatúry

ANDĚL, J. (2007). *Základy matematické statistiky.* Druhé opravené vydání. Matfyzpress, Praha. ISBN 80-7378-001-1.

DUNN, P. (2004). Occurrence and quantity of precipitation can be modelled simultaneously. *International Journal of Climatology*, **24**(10), 1231–1239.

DUPAČ, V. a HUŠKOVÁ, M. (2001). *Pravděpodobnost a matematická statistika.* Nakladatelství Karolinum, Praha. ISBN 80-246-0009-9.

JØRGENSEN, B. a PAES DE SOUZA, M. C. (1994). Fitting Tweedie's compound Poisson model to insurance claims data. *Scand. Actuar. J.*, **1**, 69–93.

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

SMYTH, G. K. a JØRGENSEN, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *Astin Bulletin*, **32**, 143–157.

WÜTHRICH, M. V. (2003). Claims reserving using Tweedie's compound Poisson model. *Astin Bulletin*, **33**(2), 331–346.

WÜTHRICH, M. V. a MERZ, M. (2008). *Stochastic Claims Reserving Methods in Insurance.* Wiley Finance, Chichester. ISBN 978-0-470-72346-3.

YOUNG, G. A. a SMITH, R. L. (2005). *Essentials of Statistical Inference.* Cambridge University Press, New York. ISBN 0-521-54866-7.

ZVÁRA, K. a ŠTĚPÁN, J. (2001). *Pravděpodobnost a matematická statistika.* Matfyzpress, Praha. ISBN 80-2240736-4.

Zoznam obrázkov