

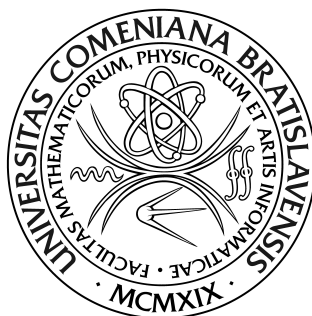
UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

**ŠTATISTICKÉ ODHADY PARAMETROV
ROZDELENÍ V POISŤOVNÍCTVE**
DIPLOMOVÁ PRÁCA

2012

BC. JANA KÁČERIKOVÁ

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



**ŠTATISTICKÉ ODHADY PARAMETROV
ROZDELENÍ V POISŤOVNÍCTVE**

DIPLOMOVÁ PRÁCA

Študijný program: Ekonomická a finančná matematika
Študijný odbor: Aplikovaná matematika 1 114
Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky
Vedúci diplomovej práce: prof. RNDr. Rastislav Potocký, PhD.

Bratislava 2012

BC. JANA KÁČERIKOVÁ



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Jana Káčeriková
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov: Štatistické odhady parametrov rozdelení v poisťovníctve

Cieľ: Štatistické odhady parametrov rozdelení v poisťovníctve

Vedúci: prof. RNDr. Rastislav Potočný, PhD.

Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky

Dátum zadania: 13.01.2011

Dátum schválenia: 14.01.2011

prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

študent

vedúci práce

Čestné prehlásenie

Čestne prehlasujem, že som diplomovú prácu vypracovala samostatne s použitím uvedenej literatúry a odbornej pomoci diplomového vedúceho.

Bratislava, 26. apríl 2012

.....

Jana Káčeriková

Pod'akovanie

Ďakujem svojmu vedúcemu diplomovej práce prof. RNDr. Rastislavovi Potockému, PhD. za odborné vedenie, pripomienky, cenné rady a ochotu prejavenu pri písaní tejto diplomovej práce.

Abstrakt

Káčeriková, Jana: *Štatistické odhady parametrov rozdelení v poisťovníctve* [Diplomová práca].
Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky.

Veduci diplomovej práce: prof. RNDr. Rastislav Potocký, PhD.

Bratislava 2012. 48 strán.

Diplomová práca sa venuje štatistickým odhadom parametrov rozdelení v poisťovníctve. V poslednom čase sa poisťovne v dôsledku rôznych prírodných katastrof čoraz častejšie stretávajú aj s nárokmi s extrémnou hodnotou. Tieto nároky popisujú rozdelenia s ťažkými chvostami. Cieľom práce je ukázať efektívnosť metódy t-score momentov pri odhadovaní parametrov rozdelení s ťažkými chvostami. Táto metóda sa javí byť vážnym konkurentom metódy maximálnej vierohodnosti, najpoužívanejšej metódy na odhad parametrov rozdelení. V práci sú spracované teoretické poznatky o daných metódach, uvedené porovnanie odhadov získaných týmito dvoma metódami a analýza senzitivnosti pre Paretovo rozdelenie. Okrem týchto dvoch metód je v práci spomenutá aj metóda najmenších štvorcov na odhad parametrov rozdelenia v životnom poistení.

Kľúčové slová: metóda t-score momentov, metóda maximálnej vierohodnosti, metóda najmenších štvorcov, odhad parametra, rozdelenie s ťažkými chvostami, Paretovo rozdelenie, analýza senzitivnosti.

Abstract

Káčeriková, Jana: *The Statistical Estimations of Parameters of The Distributions in Insurance* [Master Thesis].

Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics.

Supervisor: prof. RNDr. Rastislav Potocký, PhD.

Bratislava 2012. 48 pages.

This master thesis discusses the statistical estimations of parameters of the distributions in insurance. Recently the insurance companies due to various natural disasters are increasingly faced to the claims of extreme value. These claims are described by heavy tailed distributions. The aim of the work is to demonstrate the effectiveness of the method of t-score moments to estimate parameters of heavy tailed distributions. This method appears to be a serious competitor of the method of maximum likelihood, the most frequently used method to estimate parameters of the distributions. In the paper the theoretical knowledge of the methods, comparison of estimates obtained by these two methods and analysis of sensitivity for Pareto distribution are represented. In addition to these two methods the method of least squares to estimate parameters of the distribution in life insurance is mentioned.

Keywords: method of t-score moments, method of maximum likelihood, method of least squares, estimation of parameter, heavy tailed distribution, Pareto distribution, analysis of sensitivity.

Obsah

Zoznam tabuliek	3
Zoznam obrázkov	3
Úvod	4
1 Základné pojmy a definície	6
2 Rozdelenia v poisťovníctve	9
2.1 Rozdelenia v neživotnom poistení	9
2.1.1 Rozdelenia s ľahkými chvostami	10
2.1.2 Rozdelenia s ťažkými chvostami	11
2.2 Rozdelenia v životnom poistení	14
3 Metódy odhadu parametrov	15
3.1 Metóda maximálnej vierohodnosti	15
3.1.1 Princíp metódy	16
3.1.2 Príklady	17
3.2 Metóda najmenších štvorcov	19
3.3 Metóda momentov	22
4 Metóda <i>t-score</i> momentov	24
4.1 Funkcia <i>t-score</i>	25
4.1.1 Rozdelenia definované na $\mathbb{R}$	25
4.1.2 Rozdelenia definované na $\mathcal{X} \neq \mathbb{R}$	25
4.2 Johnson score, Johnsonova stredná hodnota a Johnsonova variancia	27
4.3 Odhady	28
4.4 Príklady	30

5	Aplikácia odhadov <i>t-score</i>	36
5.1	Porovnanie odhadov	36
5.1.1	Paretovo rozdelenie	37
5.1.2	Inverzné gamma rozdelenie	39
5.1.3	Beta rozdelenie druhého druhu	39
5.2	Odl'ahlé pozorovania	40
5.3	Kontaminované pozorovania	41
5.4	Agregované nároky	43
	Záver	45
	Literatúra	47

Zoznam tabuliek

5.1	Výsledky odhadov pre jednoduchú verziu Paretovho rozdelenia	37
5.2	Výsledky odhadov pre zovšeobecnenú verziu Paretovho rozdelenia	38
5.3	Výsledky odhadov pre inverzné gamma rozdelenie	39
5.4	Výsledky odhadov pre beta rozdelenie druhého druhu	40
5.5	Výsledky odhadov pre odl'ahlé pozorovania	41
5.6	Výsledky odhadov pre kontaminované pozorovania	42

Zoznam obrázkov

5.1	Graf odhadov parametra Paretovho rozdelenia pre rôzne miery poškodenia výberu odl'ahlými pozorovaniami	41
5.2	Graf odhadov parametra Paretovho rozdelenia pre rôzne miery poškodenia výberu kontaminovanými pozorovaniami	42
5.3	Graf hustoty sumy individuálnych nárokov	44

Úvod

Poistovníctvo je súčasťou finančného trhu a má dôležitý význam v národnom hospodárstve každej krajiny. Jeho význam nespočíva len vo veľkom finančnom obrate, ale pomáha aj udržiavať ekonomickú stabilitu podnikateľského prostredia a životnú úroveň obyvateľstva.

Jednou z úloh poisťovne je správne určenie výšky poistného. Tento proces pozostáva z niekoľkých krokov, pričom jedným z nich je odhad parametrov rozdelenia očakávaných nárokov a práve štatistickým odhadom parametrov rozdelení v poisťovníctve sa venuje táto diplomová práca.

Cieľom tejto práce je teoretické spracovanie najznámejších metód na odhad parametrov rozdelení a ukázať efektívnosť novej modernej metódy na odhad parametrov, metódy *t-score* momentov, pre triedu rozdelení s ťažkými chvostami. Našu pozornosť sústredíme na rozdelenia s ťažkými chvostami, pretože v dôsledku veľkých prírodných katastrof a nehôd sa poisťovne v poslednom čase musia vysporiadať s extrémnymi hodnotami nárokov v prípade vzniku poistnej udalosti, a tieto nároky najlepšie popisujú rozdelenia tohto typu. Štruktúra diplomovej práce pozostáva z piatich kapitol.

V prvej kapitole sú uvedené základné poznatky z teórie pravdepodobnosti a štatistiky, ktoré sú potrebné k spracovaniu danej problematiky.

Druhá kapitola sa venuje základným typom rozdelení, ktoré najlepšie popisujú skutočné hodnoty v poisťovníctve. Spomenieme rozdelenia s ľahkými chvostami, ale predovšetkým rozdelenia s ťažkými chvostami, definície ich hustôt a hodnoty základných charakteristík. Okrem toho spomenieme aj rozdelenia, ktoré modeluje počet rokov dožitia sa osoby.

V tretej kapitole je spracovaná teória ku klasickým metódam na odhad parametrov rozdelení, a to k metóde maximálnej vierohodnosti, metóde najmenších štvorcov a metóde momentov. Metódou maximálnej vierohodnosti odhadneme parametre niektorých rozdelení v neživotnom poistení. Na odhad parametrov rozdelenia v životnom poistení použijeme metódu najmenších štvorcov.

Štvrtá kapitola obsahuje odvodenie metódy *t-score* momentov, definície základných pojmov, vlastností odhadov získaných touto metódou a analytické tvary odhadov parametrov niektorých rozdelení. V porovnaní s metódou maximálnej vierohodnosti vykazuje táto metóda

robustnosť odhadov. Táto vlastnosť je veľmi dôležitá pri modelovaní nárokov s extrémnymi hodnotami. K spracovaniu tejto kapitoly použijeme predovšetkým práce Z. Fabiána [3], [4], [5] a [6].

Piata kapitola využíva teoretické poznatky z predchádzajúcich kapitol a porovnáva odhady parametrov rozdelení s ťažkými chvostami získaných metódou maximálnej vierohodnosti a metódou *t-score* momentov. Ďalej sa zaoberá efektivitou týchto dvoch metód v prípade odl'ahľých a kontaminovaných pozorovaní pre Paretovo rozdelenie. V praxi sa ukazuje, že jeden parameter tohto rozdelenia je rovný konštante jedna [10] a práve pre tento typ Paretovho rozdelenia urobíme analýzu senzitivnosti. Na záver preskúmame, aký vplyv majú tieto metódy na agregované nároky pochádzajúce z loggamma rozdelenia. Inšpirácia pre spracovanie tejto kapitoly pochádza z článkov [8] a [11].

Kapitola 1

Základné pojmy a definície

V úvodnej kapitole si zdefinujeme základné pojmy z teórie pravdepodobnosti a štatistiky, ktoré budeme využívať v nasledujúcich častiach. Ich znenie, ale aj znenie mnohých ďalších štatistických pojmov môžeme nájsť v [9].

Definícia 1.1 (Náhodná premenná)

Nech je daný pravdepodobnostný priestor $(\Omega, \mathcal{S}, P)$. Zobrazenie $X : \Omega \rightarrow R$ sa nazýva náhodná premenná, ak pre všetky $x \in R$ platí

$$\{\omega; X(\omega) < x\} \in \mathcal{S},$$

kde Ω je priestor elementárnych udalostí, $\mathcal{S}$ je σ – algebra a P je pravdepodobnosť.

Pravdepodobné správanie sa náhodnej premennej vyjadruje jej distribučná funkcia.

Definícia 1.2 (Distribučná funkcia)

Distribučná funkcia náhodnej premennej X je pre každé reálne x definovaná vzt'ahom

$$F(x) = P(X < x).$$

Teda distribučná funkcia vyjadruje pravdepodobnosť, že premenná X nadobúda hodnoty menšie ako reálne číslo x . V práci uvažujeme len spojité náhodné premenné.

Definícia 1.3 (Spojitá náhodná premenná)

Náhodná premenná X sa nazýva spojitá, ak existuje taká nezáporná integrovateľná funkcia f , že pre všetky $x \in R$ platí

$$F(x) = \int_{-\infty}^x f(t)dt.$$

Funkcia $f(x)$ sa nazýva hustota náhodnej premennej.

Nasledujúce definície sú uvedené pre spojité náhodné premenné. Medzi základné charakteristiky premennej patrí stredná hodnota a disperzia.

Definícia 1.4 (Stredná hodnota)

Stredná hodnota náhodnej premennej X s hustotou $f(x)$ je definovaná vzt'ahom

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx,$$

ak $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$. Inak hovoríme, že X nemá strednú hodnotu.

Definícia 1.5 (Disperzia)

Nech X je integrovateľná (t.j. existuje jej stredná hodnota) náhodná premenná. Ak je aj náhodná premenná $(X - E[X])^2$ integrovateľná, tak disperzia náhodnej premennej X je definovaná vzt'ahom

$$D[X] = E[(X - E[X])^2].$$

Teda disperzia vyjadruje strednú hodnotu štvorca rozdielu premennej X a jej strednej hodnoty $E[X]$. Stredná hodnota je špeciálnym prípadom skupiny parametrov, ktorú nazývame momentami.

Definícia 1.6 (Moment k -tého rádu)

Ak X^k je integrovateľná náhodná premenná, kde k je prirodzené číslo, tak moment k -tého rádu náhodnej premennej X je definovaný vzt'ahom

$$E[X^k] = \int_R x^k f(x) dx,$$

kde $f(x)$ je hustota premennej X .

Definícia 1.7 (Vytvárajúca momentová funkcia)

Vytvárajúca momentová funkcia náhodnej premennej X je definovaná vzt'ahom

$$m_x(t) = E[e^{tx}] = \int_R e^{tx} f(x) dx,$$

kde $f(x)$ je hustota premennej X .

Pri definícii hustoty premennej sa často využíva beta funkcia alebo gamma funkcia. Ich znenie, vlastnosti a vzt'ahy medzi nimi, ktoré sa využívajú pri rátaní strednej hodnoty a disperzie, môžeme nájsť v [12].

Definícia 1.8 (Beta funkcia)

Beta funkcia je definovaná vzt'ahom

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt, \quad (1.1)$$

pre $x > 0, y > 0$.

Definícia 1.9 (Gamma funkcia)

Gamma funkcia je definovaná vzt'ahom

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \quad (1.2)$$

pre $x > 0$.

Kapitola 2

Rozdelenia v poist'ovnictve

Poist'ovnictvo je odvetvie hospodárstva zaoberajúce sa poist'ovacou, zaist'ovacou a sprostredkovateľskou činnosťou v oblasti poistenia. Poistenie je právny vzťah, pri ktorom poist'ovňa preberá na seba záväzok, že poistenému (alebo poškodenému za poisteného) poskytne poistné plnenie(poistný nárok) za udalosť, ktorá vznikla v zmysle dohodnutých poistných podmienok. Tento záväzok je realizovaný za poistné. Poistenie delíme na neživotné a životné.

V oblasti neživotného poistenia potrebuje poist'ovňa pre správne určenie hodnoty poistného poznať, aké výšky poistných plnení má očakávať v prípade vzniku poistnej udalosti. V reči matematiky to znamená, že potrebuje poznať rozdelenie výšky poistných plnení. V oblasti životného poistenia má vplyv na výpočet poistného intenzita úmrtnosti.

V tejto kapitole popíšeme základné typy rozdelení, ktoré v praxi najlepšie popisujú skutočné hodnoty. Definície rozdelení sú uvedené v [1], pričom základné charakteristiky rozdelení sme odvodili.

2.1 Rozdelenia v neživotnom poistení

Oblasť neživotného poistenia je charakteristická tým, že sa tu vyskytujú aj poistné udalosti s výrazne vyššou hodnotou poistného plnenia. Náhodné premenné majú veľký rozptyl a obvykle sú zošíkmené. Dlhoročná skúsenosť ukazuje, že rozdelenie výšky celkových plnení výrazne závisí od najväčšieho plnenia, je to v prípade zemetrasenia, záplavy, veľkého požiaru atď.

V práci budeme symbolom X označovať veľkosť poistného plnenia. Z praxe je zrejmé, že pre každé X platí $X > 0$, pretože veľkosť poistného plnenia nemôže byť záporná. Nasledujúce rozdelenia túto podmienku spĺňajú. Ako najvhodnejšie rozdelenia sa javia rozdelenia

s pomerne rýchlou konvergenciou k nule a zošikmením na pravej strane. Na základe toho delíme rozdelenia na tie, ktoré sú vhodné pre malé poistné plnenia a na tie, ktoré sú vhodné pre vysoké poistné plnenia. Výška poistných plnení sa modeluje pomocou spojitých rozdelení.

2.1.1 Rozdelenia s ľahkými chvostami

Rozdelenia s ľahkými chvostami nazývame tie rozdelenia, ktoré sú vhodné pre malé poistné plnenia. Sú charakteristické tým, že pomerne rýchlo konvergujú k nule a ich vytvárajúca momentová funkcia je v nejakom kladnom okolí t konečná, teda platí nasledujúci vzťah

$$m_x(t) = \int_R e^{tx} f(x) dx < \infty.$$

Najznámejšie rozdelenie z tejto skupiny je exponenciálne rozdelenie.

Definícia 2.1 (Exponenciálne rozdelenie)

Náhodná premenná X má exponenciálne rozdelenie s parametrom λ práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = \lambda e^{-\lambda x} \quad (2.1)$$

pre $x > 0, \lambda > 0$.

Základné charakteristiky exponenciálneho rozdelenia:

stredná hodnota: $E[X] = \frac{1}{\lambda}$,

disperzia: $D[X] = \frac{1}{\lambda^2}$,

vytvárajúca momentová funkcia: $m_x(t) = \frac{\lambda}{\lambda - t}$ pre $t < \lambda$.

Exponenciálne rozdelenie patrí medzi najjednoduchšie rozdelenia z hľadiska vyjadrenia, avšak jeho nevýhodou je, že nedostatočne popisuje intervaly najnižších a najvyšších hodnôt, preto sa v praxi používa veľmi málo.

Ďalšie rozdelenia s ľahkými chvostami sú gamma rozdelenie a zmes exponenciálnych rozdelení.

Definícia 2.2 (Gamma rozdelenie)

Náhodná premenná X má gamma rozdelenie s parametrami α, γ práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\gamma x} \quad (2.2)$$

pre $x > 0, \alpha > 0, \gamma > 0$, kde $\Gamma(\alpha)$ je gamma funkcia (1.2).

Základné charakteristiky gamma rozdelenia:

stredná hodnota: $E[X] = \frac{\alpha}{\gamma}$,

disperzia: $D[X] = \frac{\alpha}{\gamma^2}$,

vytvárajúca momentová funkcia: $m_x(t) = \left(\frac{\gamma}{\gamma - t}\right)^\alpha$ pre $t < \gamma$.

Definícia 2.3 (Zmes exponenciálnych rozdelení)

Náhodná premenná X má konečnú zmes exponenciálnych rozdelení s n zložkami práve vtedy, ak jej hustota má tvar

$$f(x) = \sum_{i=1}^n p_i \lambda_i e^{-\lambda_i x}, \quad (2.3)$$

pre $x > 0$, $\lambda_i > 0$, $p_i > 0$, kde p_i je zmiešavací pomer a platí $\sum_{i=1}^n p_i = 1$.

Základné charakteristiky zmesi exponenciálnych rozdelení:

stredná hodnota: $E[X] = \sum_{i=1}^n \frac{p_i}{\lambda_i}$,

disperzia: $D[X] = 2 \sum_{i=1}^n \frac{p_i}{(\lambda_i)^2} - \left(\sum_{i=1}^n \frac{p_i}{\lambda_i} \right)^2$,

vytvárajúca momentová funkcia: $m_x(t) = \sum_{i=1}^n p_i \frac{\lambda_i}{\lambda_i - t}$ pre $t < \lambda_i$, kde $i = 1, \dots, n$.

2.1.2 Rozdelenia s ťažkými chvostami

Rozdelenia s ťažkými chvostami nazývame tie rozdelenia, ktoré sú vhodné pre vysoké poisťné plnenia. Tieto rozdelenia pomalšie konvergujú k nule ako rozdelenia s ľahkými chvostami a ich vytvárajúca momentová funkcia neexistuje. Pre všetky parametre $t > 0$ je nekonečná, teda platí nasledujúci vzťah

$$m_x(t) = \int_R e^{tx} f(x) dx = \infty.$$

Medzi najznámejšie rozdelenia s ťažkými chvostami patria Paretovo rozdelenie, lognormálne rozdelenie, loggamma rozdelenia a Weibullovo rozdelenie.

Paretovo rozdelenie má okrem modelovania výšky poisťných plnení veľké uplatnenie aj v riadení rizika, napr. pri modelovaní jednodenných výnosov a strát akcií. Používa sa v dvoch verziách, v európskej a americkej. V práci sa budeme venovať európskej verzii.

Definícia 2.4 (Paretovo rozdelenie)

Náhodná premenná X má Paretovo rozdelenie s parametrami c, a práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = c \frac{a^c}{x^{c+1}} \quad (2.4)$$

pre $x > a, a > 0, c > 0$.

Základné charakteristiky Paretovho rozdelenia:

stredná hodnota: $E[X] = \frac{ac}{c-1}$ pre $c > 1$,

disperzia: $D[X] = \frac{ca^2}{(c-1)^2(c-2)}$ pre $c > 2$.

Lognormálne rozdelenie a loggamma rozdelenie sa najčastejšie používa pri poistení proti prírodným katastrofám.

Definícia 2.5 (Lognormálne rozdelenie)

Náhodná premenná X má lognormálne rozdelenie s parametrami μ, σ práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (2.5)$$

pre $x > 0$.

Základné charakteristiky lognormálneho rozdelenia:

stredná hodnota: $E[X] = e^{\mu + \frac{1}{2}\sigma^2}$,

disperzia: $D[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$.

Definícia 2.6 (Loggamma rozdelenie)

Náhodná premenná X má loggamma rozdelenie s parametrami α, γ práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = \frac{\gamma^\alpha (\ln x)^{\alpha-1}}{x^{\gamma+1} \Gamma(\alpha)} \quad (2.6)$$

pre $x > 1, \alpha > 0, \gamma > 0$, kde $\Gamma(\alpha)$ je gamma funkcia (1.2).

Základné charakteristiky loggamma rozdelenia:

stredná hodnota: $E[X] = \left(1 - \frac{1}{\gamma}\right)^{-\alpha}$,

disperzia: $D[X] = \left(1 - \frac{2}{\gamma}\right)^{-\alpha} - \left(1 - \frac{1}{\gamma}\right)^{-2\alpha}$.

Definícia 2.7 (Inverzné gamma rozdelenie)

Náhodná premenná X má inverzné gamma rozdelenie s parametrami α, γ práve vtedy, ak jej

hustota pravdepodobnosti má tvar

$$f(x) = \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\gamma}{x}} \quad (2.7)$$

pre $x > 0, \alpha > 0, \gamma > 0$, kde $\Gamma(\alpha)$ je gamma funkcia (1.2).

Základné charakteristiky inverzného gamma rozdelenia:

stredná hodnota: $E[X] = \frac{\gamma}{\alpha - 1}$ pre $\alpha > 1$,

disperzia: $D[X] = \frac{\gamma^2}{(\alpha - 1)^2(\alpha - 2)}$ pre $\alpha > 2$.

Definícia 2.8 (Weibullovo rozdelenie)

Náhodná premenná X má Weibullovo rozdelenie s parametrami k, λ práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} e^{-\left(\frac{x}{\lambda} \right)^k} \quad (2.8)$$

pre $x > 0, \lambda > 0, k > 0$.

Základné charakteristiky Weibullovoho rozdelenia:

stredná hodnota: $E[X] = \lambda \Gamma \left(1 + \frac{1}{k} \right)$

disperzia: $D[X] = \lambda^2 \left[\Gamma \left(1 + \frac{2}{k} \right) - \Gamma \left(1 + \frac{1}{k} \right)^2 \right]$.

Medzi rozdelenia s ťažkými chvostami zaradujeme aj beta rozdelenie druhého druhu, nazývané tiež Pearsonovo rozdelenie IV. typu.

Definícia 2.9 (Beta rozdelenie druhého druhu)

Náhodná premenná X má beta rozdelenie druhého druhu s parametrami p, q práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = \frac{1}{B(p, q)} \frac{x^{p-1}}{(x+1)^{p+q}} \quad (2.9)$$

pre $x > 0, p > 0, q > 0$, kde $B(p, q)$ je beta funkcia (1.1).

Základné charakteristiky beta rozdelenia druhého druhu:

stredná hodnota: $E[X] = \frac{p}{q-1}$ pre $q > 1$,

disperzia: $D[X] = \frac{p(p+q-1)}{(q-1)^2(q-2)}$ pre $q > 2$.

2.2 Rozdelenia v životnom poistení

V životnom poistení má pri výpočte poistného dôležitý význam intenzita úmrtnosti $\mu(t)$, ktorá zodpovedá pravdepodobnosti dožitia sa t rokov a "takmer okamžitého" úmrtia. Intenzita úmrtnosti v čase t je definovaná vzťahom

$$\mu(t) = \frac{F'(t)}{1 - F(t)},$$

kde $F(t)$ je distribučná funkcia a je diferencovateľná podľa svojej premennej.

Podľa tvaru intenzity úmrtnosti vieme určiť, z akého rozdelenia je premenná t .

Rozlišujeme tieto prípady:

- nech $\mu(t) = \mu = \text{konšt.}$, potom t je opísané exponenciálnym rozdelením (2.1),
- nech $\mu(t) = at^k$, kde $a > 0, k > 0$, potom t je opísané Weibullovým rozdelením (2.8),
- nech $\mu(t) = BC^t$, kde $B > 0, C \geq 1$, potom t je opísané Gompertzovým rozdelením,
- nech $\mu(t) = A + BC^t$, kde $B > 0, A \geq B, C > 1$, potom t je opísané Makehamovým rozdelením.

Prvý prípad sa uvádza len pre porovnanie, pretože v životnom poistení nemá zmysel konštantná intenzita úmrtnosti.

Definícia 2.10 (Makehamovo rozdelenie)

Náhodná premenná X má Makehamovo rozdelenie s parametrami A, B, C práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = (A + BC^x)e^{(-Ax - \frac{B}{\ln C}(C^x - 1))} \quad (2.10)$$

pre $x > 0, B > 0, A \geq B$ a $C > 1$.

Hustota Makehamovho rozdelenia je odvodená z [7].

Kapitola 3

Metódy odhadu parametrov

V predchádzajúcej kapitole sme zadefinovali rozdelenia, ktoré najlepšie popisujú hodnoty poistných plnení. V praxi je však k dispozícii len súbor týchto poistných plnení a úlohou poisťovne je na základe týchto dát nájsť najvhodnejšie rozdelenie na modelovanie výšky poistných plnení. Prvým krokom procesu hľadania vhodného rozdelenia je výber rozdelenia na základe grafického znázornenia dát histogramom spolu s jeho základnými charakteristikami. V ďalšom kroku je potrebné určiť hodnoty parametrov daného rozdelenia a práve metódam odhadu parametrov sa venuje táto kapitola.

Odhady rozdelíme do dvoch skupín, na bodové a intervalové odhady. V tejto práci sa budeme zaoberať len metódami odhadov z prvej skupiny.

V tejto kapitole spomenieme metódu maximálnej vierohodnosti, jej princíp, vlastností odhadov a na konkrétnych rozdeleniach ukážeme, aké analytické vyjadrenia majú odhady parametrov príslušných rozdelení v neživotnom poistení získané touto metódou. Na získanie odhadov parametrov rozdelenia zo životného poistenia aplikujeme metódu najmenších štvorcov, ktorej teoretický princíp tiež zhrieme. Okrem týchto dvoch metód spomenieme aj metódu momentov. Materiál použitý k spracovaniu tejto kapitoly môžeme nájsť v [9].

3.1 Metóda maximálnej vierohodnosti

Medzi najznámejšie metódy odhadu parametrov patrí metóda maximálnej vierohodnosti. Táto metóda má široké uplatnenie a jej odhady majú dobré vlastnosti v porovnaní s inými metódami. Jej nevýhodou je citlivosť na extrémne pozorovanie, teda nevykazuje robustnosť odhadov.

3.1.1 Princíp metódy

Metóda je založená na vierohodnostnej funkcii (*likelihood function*), ktorá je definovaná pomocou hustoty.

Nech $X_1, \dots, X_n$ je náhodný výber z rozdelenia so známym tvarom hustoty pravdepodobnosti $f(x; \theta)$. Avšak vektor parametrov $\theta = (\theta_1, \dots, \theta_m)^T$ je neznámy. Vie sa len to, že patrí do známej množiny $\Theta \subset \mathbb{R}^m$ (tzv. parametrický priestor).

Potom funkciou vierohodnosti nazývame

$$L(\mathbf{x}; \theta) = L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

kde x_i je realizácia X_i .

Metóda maximálnej vierohodnosti spočíva v dosadení hodnôt parametrov $\hat{\theta}_{ML1}, \dots, \hat{\theta}_{MLm}$ za odhad neznámych hodnôt parametrov $\theta_1, \dots, \theta_m$, ktoré pri daných realizovaných hodnotách premenných maximalizujú funkciu vierohodnosti.

Definícia 3.1 (Odhad maximálnej vierohodnosti)

Odhadom maximálnej vierohodnosti vektora parametrov θ nazývame náhodný vektor $\hat{\theta}_{ML}$ taký, že platí

$$L(\mathbf{x}; \hat{\theta}_{ML}) \geq L(\mathbf{x}; \theta)$$

pre všetky $\theta \in \Theta$.

Úlohou je nájsť maximum funkcie $L(\mathbf{x}; \theta)$ v závislosti od θ . Táto funkcia je definovaná ako súčin, v tomto prípade je náročné najst' deriváciu $\frac{\partial L(\mathbf{x}; \theta)}{\partial \theta}$ na získanie maxima . Preto sa častejšie používa prirodzený logaritmus funkcie vierohodnosti, teda

$$l(\mathbf{x}; \theta) = \ln L(\mathbf{x}; \theta) = \sum_{i=1}^n \ln f(x_i; \theta).$$

Ak existujú parciálne derivácie prirodzeného logaritmu vierohodnostnej funkcie $l(\mathbf{x}; \theta)$, tak odhadovaný vektor $\hat{\theta}_{ML}$ dostaneme ako riešenie systému vierohodnostných rovníc

$$\sum_{i=1}^n \frac{\partial \ln f(x_i; \theta)}{\partial \theta_k} = 0 \quad \text{pre } k = 1, \dots, m. \quad (3.1)$$

Vlastnosti odhadu maximálnej vierohodnosti

1. konzistentný

Odhad $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ je konzistentný, ak pre $\forall \epsilon > 0$ platí

$$\lim_{n \rightarrow \infty} Pr[|\hat{\theta}_n - \theta| \geq \epsilon] = 0.$$

Táto vlastnosť hovorí o malej pravdepodobnosti veľkej chyby, ak rozsah výberu dostatočne rastie.

2. asymptoticky normálny

Odhad $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ je asymptoticky normálny, ak platí

$$\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \mathcal{N}(\theta, V)$$

kde V je kladne definitná matica.

3. asymptoticky efektívny

Odhad $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ je asymptoticky efektívny, ak platí

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V),$$

kde V je kladne definitná matic a symbolom $\xrightarrow{d}$ označujeme konvergenciu podľa distribučnej funkcie.

3.1.2 Príklady

Teraz vyjadríme analytické tvary odhadov maximálnej vierohodnosti pre niektoré rozdelenia. V prvej časti sa pozrieme na odhady rozdelení s ľahkými chvostami.

Príklad 1: Odhad parametra exponenciálneho rozdelenia.

Odhad $\hat{\lambda}_{ML}$ dostaneme ako riešenie rovnice (3.1), kde

$$l(x; \lambda) = \sum_{i=1}^n \ln(\lambda e^{-\lambda x_i}) = \sum_{i=1}^n (\ln \lambda - \lambda x_i).$$

Po dosadení do rovnice a derivovaní podľa parametra λ dostaneme

$$0 = \sum_{i=1}^n \left(\frac{1}{\lambda} - x_i \right).$$

Vyjadrením λ z predchádzajúcej rovnice dostaneme odhad pre parameter exponenciálneho rozdelenia

$$\hat{\lambda}_{ML} = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i}.$$

Príklad 2: Odhad parametra lognormálneho rozdelenia.

Odhad $\hat{\mu}_{ML}, \hat{\sigma}_{ML}^2$ dostaneme ako riešenie rovnice (3.1), kde

$$\begin{aligned} l(x; \mu, \sigma) &= \sum_{i=1}^n \ln \left(\frac{1}{\sigma x_i \sqrt{2\pi}} e^{-\frac{(\ln x_i - \mu)^2}{2\sigma^2}} \right) = \\ &= \sum_{i=1}^n \left(-\ln \sigma - \ln x_i - \ln \sqrt{2\pi} - \frac{(\ln x_i - \mu)^2}{2\sigma^2} \right). \end{aligned}$$

Po dosadení do rovnice a derivovaní podľa parametrov μ a σ dostaneme

$$\begin{aligned} 0 &= \sum_{i=1}^n \frac{(\ln x_i - \mu)}{\sigma^2} \\ 0 &= \sum_{i=1}^n \left(-\frac{1}{\sigma} + \frac{(\ln x_i - \mu)^2}{\sigma^3} \right). \end{aligned}$$

Vyjadrením μ a σ^2 z predchádzajúcich rovníc dostaneme odhady pre parametre lognormálneho rozdelenia

$$\begin{aligned} \hat{\mu}_{ML} &= \frac{1}{n} \sum_{i=1}^n \ln x_i \\ \hat{\sigma}_{ML}^2 &= \frac{1}{n} \sum_{i=1}^n (\ln x_i - \hat{\mu}_{ML})^2. \end{aligned}$$

V nasledujúcej časti sú spracované odhady Paretovho rozdelenia pre jednoduchú verziu a zovšeobecnenú verziu.

Príklad 3: Odhad parametrov Paretovho rozdelenia.

a) jednoduchá verzia

Najskôr odhadneme parameter pre jednoduchú verziu Paretovho rozdelenia, ide o rozdelenie s hustotou $f(x, c) = \frac{c}{x^{(c+1)}}$. Odhad $\hat{c}_{ML}$ dostaneme ako riešenie rovnice (3.1), kde

$$l(x; c) = \sum_{i=1}^n \ln \left(\frac{c}{x_i^{(c+1)}} \right) = \sum_{i=1}^n (\ln c - (c+1) \ln x_i).$$

Po dosadení do rovnice a derivovaní podľa parametra c dostaneme

$$0 = \sum_{i=1}^n \left(\frac{1}{c} - \ln x_i \right)$$

Vyjadrením c z predchádzajúcej rovnice dostaneme odhad pre parameter Paretovho rozdelenia

$$\hat{c}_{ML} = \frac{n}{\sum_{i=1}^n \ln x_i}.$$

b) zovšeobecnená verzia

Teraz odhadneme parametre Paretoho rozdelenia pre zovšeobecnenú verziu. Odhad $\hat{c}_{ML}$ a $\hat{a}_{ML}$ dostaneme ako riešenie rovníc (3.1), kde

$$l(x; a, c) = \sum_{i=1}^n \ln \left(\frac{ca^c}{x^{c+1}} \right) = \sum_{i=1}^n (\ln c + c \ln a - (c+1) \ln x_i).$$

Po dosadení a zderivovaní podľa parametrov a a c dostaneme sústavu rovníc

$$\sum_{i=1}^n \left(\frac{c}{a} \right) = 0$$

$$\sum_{i=1}^n \left(\frac{1}{c} + \ln a - \ln x_i \right) = 0.$$

Môžeme si všimnúť, že z prvej rovnice vychádza odhad pre parameter $\hat{c}_{ML} = 0$, toto riešenie nie je vyhovujúce (po dosadení dostaneme nulovú hustotu).

Pri hľadaní odhadov parametrov sa pozrieme na funkciu

$$l(x; a, c) = \sum_{i=1}^n (\ln c + c \ln a - (c+1) \ln x_i),$$

ktorá je rastúca v bode a , t.j. väčšia hodnota parametra a spôsobí väčšiu hodnotu prirodzeného logaritmu virohodnostnej funkcie $l(x; a, c)$. Touto úvahou dostaneme odhad parametra a

$$\hat{a}_{ML} = \min_i x_i. \quad (3.2)$$

Vyjadrením parametra c z druhej rovnice dostaneme jeho odhad

$$\hat{c}_{ML} = \frac{n}{\sum_{i=1}^n (\ln x_i - \ln \hat{a}_{ML})}. \quad (3.3)$$

Nevýhodou metódy maximálnej virohodnosti pre zložité tvary hustoty niektorých rozdelení je zložitosť analytického vyjadrenia odhadov zo systému (3.1). Preto sú v mnohých programovacích prostrediach naprogramované funkcie na odhad parametrov získaných pomocou metódy maximálnej virohodnosti. V 5. kapitole budeme využívať funkcie *betafit* a *gamfit*, ktoré sú súčasťou *MatLab-u*.

3.2 Metóda najmenších štvorcov

Táto metóda sa používa na odhad parametrov modelu

$$Y = F\theta + \epsilon,$$

kde Y je n -rozmerný vektor pozorovaní, F je matica rozmeru $n \times m$, ktorá obsahuje hodnoty vysvetľujúcich premenných, θ je m -rozmerný vektor parametrov a ϵ je n -rozmerný vektor chýb.

Je založená na minimalizácii rozdielu

$$\sum_{i=1}^n \left(Y_i - \sum_{j=1}^m F_{ij} \theta_j \right)^2,$$

vzhľadom na θ . Predchádzajúci zápis môžeme vektorovo zapísať ako

$$(Y - F\theta)^T(Y - F\theta).$$

Definícia 3.2 (Odhad najmenších štvorcov)

Úloha

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^m} \sum_{i=1}^n [y_i - F_i \theta]^2$$

má pre $n > m$ a pre danú maticu F takú, že existuje $(F^T F)^{-1}$, jediné riešenie, ktoré nazývame odhadom najmenších štvorcov,

$$\hat{\theta} = (F^T F)^{-1} F^T Y,$$

kde Y je n -rozmerný vektor pozorovaní a F_i je i -tý riadok matice F .

Teraz sa pozrieme na odhady parametrov získané metódou najmenších štvorcov. Táto metóda sa ukazuje byť efektívna v prípade odhadov parametrov Makehamovho rozdelenia v životnom poistení. Budeme uvažovať prípad Makehamovho rozdelenia (2.10), kde položíme $C = e^\alpha$. Potom distribučná funkcia a funkcia hustoty má tvar

$$F(t) = 1 - \exp \left[-At - \frac{B}{\alpha} (e^{\alpha t} - 1) \right]$$

$$f(t) = (A + B e^{\alpha t}) \exp \left[-At - \frac{B}{\alpha} (e^{\alpha t} - 1) \right], \quad \alpha > 0.$$

V článku [7] je navrhnutá minimalizačná funkcia v tvare

$$\sum_{i=1}^n [\ln(1 - F(t_i; \theta)) - \ln E[1 - F(t_i; \theta)]]^2.$$

Pre zjednodušenie zápisu použijeme nasledujúce substitúcie

$$H(t_i; \theta) = -\ln(1 - F(t_i; \theta)), \quad \beta_i = -\ln(1 - E[F(t_i; \theta)]),$$

teda minimalizačnú funkciu môžeme prepísať do tvaru

$$\sum_{i=1}^n [H(t_i; \theta) - \beta_i]^2. \tag{3.4}$$

Makehamovo rozdelenie má tri parametre A, B, C .

V tejto práci rozoberieme 3 prípady :

- A je parameter a B, C sú známe
- B je parameter a A, C sú známe
- A, B sú parametre a C je známe.

V ostatných prípadoch je postup odhadu parametrov analogický. Tieto odhady a aj ďalšie prípady môže čitateľ nájsť v [7].

1. A je parameter a B, C sú známe

Je zrejmé, že minimalizačná funkcia (3.4) dosiahne svoje minimum v prípade, ak jej prvá derivácia bude nulová. Teda $H(t_i; \theta) = \beta_i$, v maticovom zápise dostaneme $\beta = X\theta$, kde

$$\beta = \begin{pmatrix} \beta_1 - \frac{B}{\alpha}(e^{\alpha t_1} - 1) \\ \beta_2 - \frac{B}{\alpha}(e^{\alpha t_2} - 1) \\ \vdots \\ \beta_n - \frac{B}{\alpha}(e^{\alpha t_n} - 1) \end{pmatrix}, \quad X = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix}, \quad \theta = A.$$

Z definície pre odhad najmenších štvorcov vyplýva, že $\hat{\theta} = (X^T X)^{-1} X^T \beta$. Po dosadení dostaneme

$$\hat{A} = \frac{\sum_{i=1}^n [\beta_i - (\frac{B}{\alpha})(e^{\alpha t_i} - 1)] t_i}{\sum_{i=1}^n t_i^2},$$

kde $\beta_i = -\ln \left[1 - \frac{i}{n+1} \right]$, $i = 1, \dots, n$.

2. B je parameter a A, C sú známe

Opäť ako v prvom prípade uvažujeme minimalizačnú funkciu (3.4), a riešenie hľadáme v tvare $\hat{\theta} = (X^T X)^{-1} X^T \beta$, kde

$$\beta = \begin{pmatrix} \beta_1 - At_1 \\ \beta_2 - At_2 \\ \vdots \\ \beta_n - At_n \end{pmatrix}, \quad X = \begin{pmatrix} \frac{1}{\alpha}(e^{\alpha t_1} - 1) \\ \frac{1}{\alpha}(e^{\alpha t_2} - 1) \\ \vdots \\ \frac{1}{\alpha}(e^{\alpha t_n} - 1) \end{pmatrix}, \quad \theta = B.$$

Po dosadení dostaneme tvar pre odhad parametra

$$\hat{B} = \frac{\sum_{i=1}^n \alpha(\beta_i - At_i)(e^{\alpha t_i} - 1)}{\sum_{i=1}^n (e^{\alpha t_i} - 1)^2},$$

kde $\beta_i = -\ln \left[1 - \frac{i}{n+1} \right]$, $i = 1, \dots, n$.

3. A, B sú parametre a C je známe

V tomto prípade prepíšeme minimalizačnú funkciu (3.4) do tvaru

$$L(A, \alpha) = \sum_{i=1}^n \left[At_i + \frac{B}{\alpha}(e^{\alpha t_i} - 1) - \beta_i \right]^2 \quad (3.5)$$

Odhady parametrov dostaneme ako riešenie sústavy parciálnych derivácií funkcie $L(A, \alpha)$ podľa príslušných parametrov

$$\begin{aligned} \frac{\partial L(A, B)}{\partial A} &= 2 \sum_{i=1}^n \left[At_i + \frac{B}{\alpha}(e^{\alpha t_i} - 1) - \beta_i \right] t_i = 0 \\ \frac{\partial L(A, B)}{\partial B} &= 2 \sum_{i=1}^n \left[At_i + \frac{B}{\alpha}(e^{\alpha t_i} - 1) - \beta_i \right] \frac{e^{\alpha t_i} - 1}{\alpha} = 0. \end{aligned}$$

Navyše musia spĺňať podmienku pre minimum

$$\frac{\partial^2 L(A, B)}{\partial A^2} > 0, \quad \frac{\partial^2 L(A, B)}{\partial A \partial B} > 0, \quad \frac{\partial^2 L(A, B)}{\partial B^2} > 0,$$

po zderivovaní dostaneme

$$\begin{aligned} \frac{\partial^2 L(A, B)}{\partial A^2} &= 2 \sum_{i=1}^n t_i^2, & \frac{\partial^2 L(A, B)}{\partial A \partial B} &= 2 \sum_{i=1}^n \frac{t_i(e^{\alpha t_i} - 1)}{\alpha}, \\ \frac{\partial^2 L(A, B)}{\partial B^2} &= 2 \sum_{i=1}^n \left(\frac{e^{\alpha t_i} - 1}{\alpha} \right)^2. \end{aligned}$$

Pri tomto prípade sme uviedli postup, sústavy rovníc, ktoré určujú riešenie, avšak analytické tvary odhadov sú veľmi zložité a náročné na vyjadrenie.

3.3 Metóda momentov

Pre niektoré rozdelenia je metóda maximálnej vierohodnosti numericky náročná, vtedy sa ukazuje metóda momentov ako vhodná voľba. Patrí medzi najstaršie metódy a je jednoduchá. Odhady získané touto metódou nemajú dobré vlastnosti, preto sa zvyčajne používajú len ako počiatočné aproximácie. Nevýhodou tejto metódy je, že ju nemôžeme použiť v prípade rozdelenia, ktoré nemá konečné momenty.

Jej princíp spočíva v nahradení základných charakteristík výberovými charakteristikami. Nech $X_1, \dots, X_n$ je náhodný výber z rozdelenia, ktoré závisí od $m, m \geq 1$ neznámych parametrov $\theta_1, \dots, \theta_m$. Nech pre každý parameter θ existujú momenty $m_k(\theta) = E(X_i^k)$, $k = 1, \dots, m$. Nech výberové momenty sú definované vzťahom

$$m_k^V(\theta) = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Potom odhady parametrov pomocou metódy momentov získame ako riešenie rovníc

$$m_k(\hat{\theta}) = m_k^V.$$

Ak prvých m rovníc jednoznačne neurčí odhad, môžeme pridať ďalšie, ale len v prípade, ak existujú ich príslušné momenty.

Kapitola 4

Metóda *t-score* momentov

V predchádzajúcej kapitole sme sa venovali najpoužívanejším metódam na odhad parametrov rozdelení. Spomenuli sme metódu maximálnej vierohodnosti, metódu najmenších štvorcov a metódu momentov.

V tejto kapitole predstavíme modernú metódu používanú na odhad parametrov niektorých rozdelení s ťažkými chvostami. Autorom tejto metódy je N.L. Johnson a jej vznik sa viaže k roku 1949. V roku 2001 Z. Fabián túto metódu zovšeobecnil .

Najväčšia výhoda tejto metódy je jej konkurencieschopnosť s najpoužívanejšou metódou na odhad parametrov rozdelení, a to metódou maximálnej vierohodnosti.

Základným pojmom kapitoly je *Johnson score*. Pomocou neho zadefinujeme *Johnsonovu strednú hodnotu* a *Johnsonovu varianciu*, čím ukážeme význam a dôležitosť tohto pojmu, pretože tieto dve charakteristiky dostatočne popisujú rozdelenie. Hlavnou zložkou *Johnson score* je funkcia *t-score*¹.

Táto metóda patrí medzi momentové metódy, rozdiel je v nahradení momentov premennej momentami funkcie *score*. V tomto je aj ďalšia výhoda tejto metódy, pretože nezávisí na existencii momentov premennej, preto je vhodná pre rozdelenia s ťažkými chvostami, pre ktoré momenty neexistujú alebo existujú len pre určité hodnoty parametrov.

V tejto kapitole spracujeme teóriu k metóde *t-score* momentov, zadefinujeme základné pojmy, vyslovíme vlastností odhadov získaných touto metódou a na konkrétnych príkladoch ukážeme, aké analytické vyjadrenia majú odhady parametrov príslušných rozdelení. Teória je spracovaná z [3], [4], [5] a [6].

¹V literatúre sa môžeme s funkciou *t-score* stretnúť aj pod názvom funkcia *core*.

4.1 Funkcia *t-score*

Funkcia *t-score* je zovšeobecnením funkcie *score*. Zovšeobecnenie sa týka použitia funkcie aj pre rozdelenia definované na $\mathcal{X} \neq \mathbb{R}$.

4.1.1 Rozdelenia definované na $\mathbb{R}$

V tejto časti kapitoly zadefinujeme funkciu *score*.

Nech G je rozdelenie náhodnej premennej definovanej na $\mathbb{R}$, nech g je jeho hustota spojitě diferencovateľná podľa svojej premennej a regulárna v zmysle

$$\int_{-\infty}^{\infty} \frac{(g'(y))^2}{g(y)} dx < \infty.$$

Funkcia *score* pre rozdelenie G je definovaná vzťahom

$$Q(y) = -\frac{g'(y)}{g(y)}, \quad (4.1)$$

pričom spĺňa nasledujúce dve vlastnosti

$$E(Q) = 0, \quad E(Q^2) < \infty.$$

Nech $\Theta = \mathbb{R} \times \Theta^{m-1}$ je parametrický priestor a $\theta = (\mu, \tilde{\theta})$ je vektor parametrov, kde $\mu \in \mathbb{R}$ sa nazýva lokalizačný parameter a je to najdôležitejší parameter rozdelenia G (hovorí o posunutí grafu hustoty po x-ovej osi).

Nech hustota rozdelenia G_μ je $g_\mu(y) = g(y - \mu)$. Potom jeho funkcia *score* má tvar

$$Q_\mu(y) = -\frac{1}{g(y - \mu)} \frac{dg(y - \mu)}{dy} = \frac{\partial}{\partial \mu} \ln g_\mu(y).$$

Môžeme si všimnúť, že je rovná funkcii maximálnej vierohodnosti pre parameter μ .

Nevýhodnou funkcie *score* je, že uvažuje len rozdelenie definované na $\mathbb{R}$.

4.1.2 Rozdelenia definované na $\mathcal{X} \neq \mathbb{R}$

Teraz sa budeme zaoberať rozdeleniami F náhodnej premennej definovanej na $\mathcal{X} = (a, b)$. Rozdelenie definované na $\mathcal{X} = \mathbb{R}$ budeme považovať za *prototyp* rozdelenia definovanom na $\mathcal{X} \neq \mathbb{R}$. Nech $\eta : \mathcal{X} \rightarrow \mathbb{R}$ je ľubovoľné spojitě, rastúce zobrazenie, pre ktoré platí $\eta^{-1} : \mathbb{R} \rightarrow \mathcal{X}$. Uvažujme náhodnú premennú Y z rozdelenia G .

Nech náhodná premenná $X = \eta^{-1}(Y)$ je z rozdelenia F . Potom rozdelenie F má distribučnú funkciu

$$F(x) = G(\eta(x)) \quad (4.2)$$

a hustotu

$$f(x) = g(\eta(x))\eta'(x), \quad (4.3)$$

kde $\eta'(x) = d\eta(x)/dx$.

Zaujímavá vlastnosť rozdelenia F je analogické vyjadrenie funkcie *score* ako transformovaná funkcia *score* svojho prototypu

$$T(x) = Q(\eta(x)). \quad (4.4)$$

Funkcia $T(x)$ sa nazýva funkcie *t-score* rozdelenia F . Použitím substitúcie $y = \eta(x)$ a vzt'ahu (4.3) máme

$$T(x) = \frac{1}{g(y)} \frac{d}{dy} (-g(y)) = \frac{\eta'(x)}{f(x)} \frac{d}{dx} \left(-\frac{f(x)}{\eta'(x)} \right) \left(\frac{dy}{dx} \right)^{-1}.$$

Dostali sme tvar pre funkciu *t-score* ľubovoľného rozdelenia bez ohľadu na prototyp, teda

$$T(x) = \frac{1}{f(x)} \frac{1}{dx} \left(-\frac{1}{\eta'(x)} f(x) \right) \quad (4.5)$$

Zobrazenie $\eta(x)$ sa nazýva *Johnsonova transformácia* a jeho voľba $\eta(x)$ závisí od intervalu, na ktorom je rozdelenie F definované. Najčastejšie prípady sú zaznamenané v nasledujúcej schéme

$$\eta(x) = \begin{cases} x, & (a, b) = \mathbb{R}; \\ \ln(x - a), & -\infty < a < b = \infty; \\ \ln\left(\frac{x - a}{b - x}\right), & -\infty < a < b < \infty; \\ \ln(b - x), & -\infty = a < b < \infty. \end{cases} \quad (4.6)$$

Pre niektoré rozdelenie sa používa špeciálny typ *Johnsonovej transformácie*. Pre loggamma rozdelenie sa volí $\eta(x) = \ln \ln x$ a pre rozdelenia, ktorých hustota je definovaná pomocou goniometrických funkcií na intervale $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, sa volí $\eta(x) = \tan(x)$. Ako prvý sa hľadáním vhodnej transformácie zaoberal Kovanic, avšak použil neštatistický model, preto sa touto transformáciou nebudeme zaoberať. Viac sa čitateľ dozvie z [5].

4.2 Johnson score, Johnsonova stredná hodnota a Johnsonova variancia

V tejto časti zadefinujeme nové pojmy súvisiace s metódou t -score momentov.

Definícia 4.1

Nech F je rozdelenie definované na $\mathcal{X} \subseteq \mathbb{R}$ a f je hustota spojitě diferencovateľná podľa svojej premennej. Nech zobrazenie $\eta : \mathcal{X} \rightarrow \mathbb{R}$ je dané vzťahom (4.6), nech $T(x)$ je funkcia t -score podľa (4.5) a riešenie x^* rovnice

$$T(x) = 0 \quad (4.7)$$

je jediné. Funkcia

$$S(x) = \eta'(x^*)T(x) \quad (4.8)$$

sa nazýva *Johnson score* rozdelenia F .

Môžeme si všimnúť, že ak je rozdelenie definované na $\mathbb{R}$, tak volíme $\eta(x) = x$, potom $\eta'(x) = 1$ a funkcia *Johnson score* (4.8) sa zhoduje s funkciou *score* (4.1).

Definícia 4.2

Nech k je prirodzené číslo, nech X je náhodná premenná z rozdelenia F a S je *Johnson score*. Potom k -tý moment je definovaný

$$E[S^k] = \int_{\mathcal{X}} S^k(x) dF(x). \quad (4.9)$$

V práci sa budeme zaoberať len prvými dvoma momentami.

Tvrdenie 4.1. Nech F rozdelenie s *Johnson score* S má prototyp rozdelenie G s funkciou *score* Q . Potom $E[S^k] = (\eta'(x^*))^k E[Q^k]$

Dôkaz :

Zo vzťahu (4.8) môžeme strednú hodnotu vyjadriť ako $E[S^k] = (\eta'(x^*))^k E[T^k]$. Ešte musíme ukázať, že platí $E[T^k] = E[Q^k]$. Použitím vzťahu (4.4), (4.3) a (4.1) dostaneme

$$\begin{aligned} E[T^k] &= \int_a^b T^k(x) f(x) dx = \int_a^b Q^k(\eta(x)) g(\eta(x)) \eta'(x) dx = \\ &= \int_{-\infty}^{\infty} Q^k(y) g(y) dy = E[Q^k]. \quad \square \end{aligned}$$

Keďže $\eta'(x) > 0$, tak z tvrdenia 4.1 a vlastností funkcie *score* vyplývajú nasledujúce dve vlastnosti pre *Johnson score* momenty

$$E[S] = 0, \quad E[S^2] < \infty.$$

Hodnota riešenia x^* rovnice (4.7) sa nazýva *Johnsonova stredná hodnota* a hodnota $\omega^2 = \frac{1}{E[S^2]}$ sa nazýva *Johnsonova variancia* rozdelenia F .

Tvrdenie 4.2. Pre rozdelenia definované na $\mathcal{X} = (0, \infty)$ platí $\omega^2 = \frac{(x^*)^2}{E[T^2]}$.

Dôkaz :

Z tvrdenia 4.1 máme $E[S^2] = (\eta'(x^*))^2 E[T^2]$ a z (4.6) vyplýva, že $\eta'(x^*) = \frac{1}{x^*}$. □

Toto tvrdenie budeme využívať pri rátaní *Johnsonovej variancii*, pretože v práci uvažujeme rozdelenia definované na $\mathcal{X} = (0, \infty)$.

Rozdelenia definované na $\mathcal{X} \neq \mathbb{R}$ sa delia na dve skupiny:

1. Do prvej skupiny patria rozdelenia s lokalizačným parametrom μ . Tieto rozdelenia majú parameter

$$t = \eta^{-1}(\mu),$$

ktorý nazývame *Johnsonov parameter*. Jeho hodnota je rovná *Johnsonovej strednej hodnote*. Pričom funkcia $S(x; t) = \eta'(t)T(x; t)$ je rovná funkcii maximálnej vierohodnosti $l(x; t) = \frac{\partial}{\partial t} \ln f(x; t)$ pre parameter t . Medzi tieto rozdelenia patrí, napr. exponenciálne rozdelenie.

2. Do druhej skupiny patria rozdelenia bez *Johnsonovho parametra*. Medzi tieto rozdelenia patrí, napr. Paretovo rozdelenie, inverzné gamma rozdelenie a iné.

4.3 Odhady

Nech $X_1, \dots, X_n$ je náhodný výber z rozdelenia $F(x, \theta)$ so známym tvarom hustoty pravdepodobnosti $f(x, \theta)$, kde θ je neznámy vektor parametrov, $\theta \in \Theta, \Theta \subseteq \mathbb{R}^m$ a $x_1, \dots, x_n$ sú získané realizácie.

Metóda *t-score* momentov patrí medzi momentové metódy. Od klasickej metódy sa odlišuje

tým, že namiesto momentov daného rozdelenia používa *Johnson score* momenty definované v definícií 4.2, teda dostaneme systém rovníc

$$\frac{1}{n} \sum_{i=1}^n S^k(x_i, \theta) = E[S^k(\theta)] \quad k = 1, \dots, n. \quad (4.10)$$

Odhady $\hat{\theta}_n$ z (4.10) sa nazývajú *t-score* odhady.

Vlastnosti *t-score* odhadu

1. konzistentný
2. asymptoticky normálny
3. asymptoticky efektívny
4. robustný (v prípade ohraničenosti funkcie *t-score*)

Zvyčajne prvá alebo prvé dve rovnice systému (4.10) dávajú konkrétne výberové odhady oboch *Johnsonových charakteristík*. Nech teda odhad $\hat{x}_n^*$ sa nazýva *Johnsonova výberová stredná hodnota* a odhad $\hat{\omega}_n^2$ sa nazýva *Johnsonova výberová disperzia*.

Môžeme si všimnúť, že hodnota $\eta'(x^*)$ vo výraze (4.8) je konštanta, tak v rovniciach (4.10) môžeme namiesto $S(x, \theta)$ písať $T(x, \theta)$. Teda systém (4.10) prepíšeme do zjednodušeného tvaru

$$\frac{1}{n} \sum_{i=1}^n T^k(x_i, \theta) = E[T^k(\theta)] \quad k = 1, \dots, n. \quad (4.11)$$

Ďalším dôsledkom predchádzajúcej teórie je úprava prvej rovnice systému (4.11)

$$\sum_{i=1}^n T(x_i, \theta) = 0 \quad k = 1, \dots, n. \quad (4.12)$$

Navyše, ak $x^*(\theta)$ je vyjadrené ako funkcia jedného parametra alebo funkcia pomeru dvoch parametrov, tak prvú rovnicu z (4.10) môžeme prepísať do tvaru

$$\sum_{i=1}^n S(x_i; x^*) = 0, \quad (4.13)$$

čím dostaneme rovnicu pre odhad *Johnsonovej strednej hodnoty*. Odhady založené na *t-score* nezávisia na existencií momentov, preto sú zaujímavé najmä pre rozdelenia s ťažkými chvostami.

4.4 Príklady

V tejto časti vyjadríme analytické tvary odhadov parametrov najpoužívanejších rozdelení v neživotnom poist'ovníctve pomocou *t-score* momentov.

Príklad 1: Odhad parametra exponenciálneho rozdelenia.

Z definície exponenciálneho rozdelenia máme $\mathbb{X} = (0, \infty)$, teda na základe schémy (4.6) volíme $\eta = \ln x$. Potom podľa vzťahu (4.5) *t-score* pre exponenciálne rozdelenie má tvar

$$T(x) = \lambda x - 1.$$

Prislúchajúce charakteristiky sú

$$S(x) = \lambda(\lambda x - 1), \quad x^* = \frac{1}{\lambda}, \quad \omega^2 = \frac{1}{\lambda^2}.$$

Z (4.13) pre odhad parametra λ dostaneme rovnicu

$$\sum_{i=1}^n (\lambda x_i - 1) = 0.$$

Odhad parametra λ pre exponenciálne rozdelenie má tvar

$$\hat{\lambda}_{TSCORE} = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i}.$$

Môžeme si všimnúť, že odhad pre parameter exponenciálneho rozdelenia získaný pomocou metódy maximálnej vierohodnosti a metódou *t-score* momentov je rovnaký.

Príklad 2: Odhad parametrov lognormálneho rozdelenia.

Z definície lognormálneho rozdelenia máme $\mathbb{X} = (0, \infty)$, teda na základe schémy (4.6) volíme $\eta = \ln x$. Potom podľa vzťahu (4.5) *t-score* pre lognormálne rozdelenie má tvar

$$T(x) = \frac{\ln x - \mu}{\sigma^2}.$$

Prislúchajúce charakteristiky sú

$$S(x) = \frac{1}{e^\mu} \frac{\ln x - \mu}{\sigma^2}, \quad x^* = e^\mu, \quad \omega^2 = \frac{\sigma^2}{e^{-2\mu}}.$$

Z (4.13) pre odhad parametrov μ a σ^2 dostaneme sústavu rovníc

$$\sum_{i=1}^n \frac{\ln x_i - \mu}{\sigma^2} = 0$$

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{\ln x_i - \mu}{\sigma^2} \right)^2 = \frac{1}{\sigma^2}.$$

Vyjadrením parametrov μ a σ^2 z predchádzajúcich dvoch rovníc dostaneme odhady parametrov lognormálneho rozdelenia

$$\begin{aligned} \hat{\mu}_{TSCORE} &= \frac{1}{n} \sum_{i=1}^n \ln x_i \\ \hat{\sigma}_{TSCORE}^2 &= \frac{1}{n} \sum_{i=1}^n (\ln x_i - \hat{\mu}_{TSCORE})^2. \end{aligned}$$

Odhady parametrov lognormálneho rozdelenia získané pomocou metódy maximálnej viero-
hodnosti a metódou *t-score* momentov sa tiež zhodujú.

Príklad 3: Odhad parametrov loggamma rozdelenia.

Pre loggamma rozdelenie volíme špeciálny typ Johnsonovej transformácie, a to $\eta = \ln \ln x$.
Potom podľa vzťahu (4.5) *t-score* pre loggamma rozdelenie má tvar

$$T(x) = \gamma \ln x - \alpha.$$

Prislúchajúce charakteristiky sú

$$S(x) = \frac{\gamma}{\alpha e^{\frac{\alpha}{\gamma}}} (\gamma \ln x - \alpha), \quad x^* = e^{\frac{\alpha}{\gamma}}, \quad \omega^2 = \frac{2}{\gamma} e^{\frac{2\alpha}{\gamma}}.$$

Z (4.13) pre odhad parametrov γ a α dostaneme sústavu rovníc

$$\begin{aligned} \sum_{i=1}^n (\gamma \ln x_i - \alpha) &= 0 \\ \frac{1}{n} \sum_{i=1}^n (\gamma \ln x_i - \alpha)^2 &= \alpha. \end{aligned}$$

Po vyjadrení α z prvej rovnice a dosadení do druhej rovnice dostaneme výraz

$$0 = \gamma \left[\gamma \sum_{i=1}^n (\ln x_i)^2 - \frac{\gamma}{n} \left(\sum_{i=1}^n \ln x_i \right)^2 - \sum_{i=1}^n \ln x_i \right].$$

Z definície loggamma rozdelenia riešenie $\gamma = 0$ vylúčime, teda pre odhad parametra γ máme
jediné riešenie. Odhady parametrov loggamma rozdelenie majú tvar

$$\hat{\gamma}_{TSCORE} = \frac{\sum_{i=1}^n \ln x_i}{\sum_{i=1}^n (\ln x_i)^2 - \frac{1}{n} \left(\sum_{i=1}^n \ln x_i \right)^2} \quad (4.14)$$

$$\hat{\alpha}_{TSCORE} = \hat{\gamma}_{TSCORE} \frac{\sum_{i=1}^n \ln x_i}{n}. \quad (4.15)$$

Medzi najčastejšie uvádzané rozdelenia, ktorých odhady sú vyjadrené pomocou metódy *t-score* momentov, patria Paretovo rozdelenie, inverzné gamma rozdelenie a beta rozdelenie druhého druhu.

Pri vyjadrovaní odhadov parametrov rozdelení metódou *t-score* momentov sa vo všeobecnosti používa nasledujúci postup. V prvom kroku sa odhadne $\hat{x}^*$ z prvej rovnice systému (4.10), pričom x^* sa dá napísať ako funkcia jedného parametra alebo podiel dvoch parametrov. Ďalším krokom je vyjadrenie jedného parametra z druhej rovnice systému (4.10) a nakoniec zo vzťahu pre x^* dostaneme odhad pre druhý parameter daného rozdelenia.

Príklad 4: Odhad parametrov Paretovho rozdelenia.

a) jednoduchá verzia

V praxi sa často využíva jednoparametrové Paretovo rozdelenie. Ide o Paretovo rozdelenie, kedy parameter $a = 1$, teda hustota tohto rozdelenia má tvar $f(x; c) = \frac{c}{x^{c+1}}$. Podľa schémy (4.6) zvolíme $\eta(x) = \ln(x - 1)$. Potom podľa (4.5) *t-score* má tvar

$$T(x) = c - \frac{1+c}{x}$$

Príslušné charakteristiky sú

$$S(x) = \frac{c}{c+1} \left(c - \frac{1+c}{x} \right), \quad x^* = \frac{1+c}{c}, \quad \omega^2 = \frac{(c+1)^2(c+2)}{c^3}.$$

Z (4.10) dostaneme rovnicu pre odhad parametra c

$$\sum_{i=1}^n \left(c - \frac{1+c}{x_i} \right) = 0.$$

Z predchádzajúcej rovnice vyjadríme odhad pre parameter Paretovho rozdelenia

$$\hat{c}_{TSCORE} = \frac{1}{\bar{x}_H - 1}, \quad (4.16)$$

kde $\bar{x}_H$ je harmonický priemer definovaný vzťahom

$$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}. \quad (4.17)$$

b) zovšeobecnená verzia

Teraz vyjadríme odhady parametrov Paretovho rozdelenia s hustotou (2.4). Z definície Paretovho rozdelenia máme $\mathbb{X} = (a, \infty)$, teda podľa schémy (4.6) volíme transformáciu $\eta(x) = \ln(x - a)$. Podľa vzťahu (4.5) *t-score* pre Paretovo rozdelenie má tvar

$$T(x) = c - \frac{a(c+1)}{x}.$$

Príslušné charakteristiky sú

$$S(x) = \frac{c}{a(c+1)} \left(c - \frac{a(c+1)}{x} \right), \quad x^* = \frac{a(c+1)}{c}, \quad \omega^2 = \frac{a^2(c+1)^2(c+2)}{c^3}.$$

Z (4.10) dostaneme sústavu rovníc pre odhad parametrov a a c

$$\sum_{i=1}^n \left(c - \frac{a(c+1)}{x_i} \right) = 0$$

$$\frac{1}{n} \sum_{i=1}^n \left(c - \frac{a(c+1)}{x_i} \right)^2 = \frac{c}{c+2}.$$

Úpravou prvej rovnice dostaneme odhad pre x^*

$$0 = \sum_{i=1}^n \left(c - \frac{a(c+1)}{x_i} \right) = \sum_{i=1}^n c \left(1 - \frac{a + \frac{a}{c}}{x_i} \right) = \sum_{i=1}^n c \left(1 - \frac{x^*}{x_i} \right),$$

teda $\hat{x}^* = \bar{x}_H$, kde $\bar{x}_H$ je harmonický priemer (4.17). Pre zjednodušenie výpočtu položíme substitúciu $A = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\bar{x}_H}{x_i} \right)^2$. Po dosadení do druhej rovnice dostaneme kvadratickú rovnicu

$$c^2 A + 2cA - 1 = 0.$$

Z definície Paretovho rozdelenia vieme, že $c > 0$, teda máme jediné riešenie pre odhad parametra c Paretovho rozdelenia

$$\hat{c}_{TSCORE} = -1 + \sqrt{1 + \frac{1}{A}}. \quad (4.18)$$

Odhad parametra a Paretovho rozdelenia má tvar

$$\hat{a}_{TSCORE} = \frac{\hat{c}_{TSCORE}}{\hat{c}_{TSCORE} + 1} \bar{x}_H. \quad (4.19)$$

Príklad 5: Odhad parametrov beta rozdelenia druhého druhu.

Na základe definície rozdelenia a podľa (4.6) volíme $\eta = \ln x$. Potom podľa vzťahu (4.5) *t-score* pre beta rozdelenie druhého druhu má tvar

$$T(x) = \frac{qx - p}{x + 1}.$$

Prislúchajúce charakteristiky sú

$$S(x) = \frac{q}{p} \left(\frac{qx - p}{x + 1} \right), \quad x^* = \frac{p}{q}, \quad \omega^2 = \frac{p(p + q + 1)}{q^3}.$$

Z (4.13) pre odhad parametrov p a c dostaneme sústavu rovníc

$$\sum_{i=1}^n \left(\frac{qx_i - p}{x_i + 1} \right) = 0$$

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{qx_i - p}{x_i + 1} \right)^2 = \frac{pq}{1 + p + q}.$$

Po dosadení x^* do prvej rovnice a jej úpravou

$$\sum_{i=1}^n q \left(\frac{x_i - \frac{p}{q}}{x_i + 1} \right) = \sum_{i=1}^n q \left(\frac{x_i - x^*}{x_i + 1} \right),$$

dostaneme tvar pre odhad x^*

$$\hat{x}^* = \frac{\sum_{i=1}^n \frac{x_i}{x_i + 1}}{\sum_{i=1}^n \frac{1}{x_i + 1}}.$$

Úpravou druhej rovnice dostaneme vzťah

$$\frac{1}{n} \sum_{i=1}^n q^2 \left(\frac{x_i - \hat{x}^*}{x_i + 1} \right)^2 = \frac{pq}{1 + p + q}$$

$$\frac{q^2(1 + p + q)}{pq} = \frac{n}{\sum_{i=1}^n \left(\frac{x_i - \hat{x}^*}{x_i + 1} \right)^2}$$

Pre zjednodušenie výpočtu položíme substitúciu $A = \frac{n}{\sum_{i=1}^n \left(\frac{x_i - \hat{x}^*}{x_i + 1} \right)^2}$, čím po jej dosadení a

úpravach predchádzajúceho vzťahu dostaneme $\frac{1}{\hat{x}^*} + q + \frac{q}{\hat{x}^*} = A$. Teda odhady majú tvar

$$\hat{q}_{TSCORE} = \frac{A\hat{x}^* - 1}{\hat{x}^* + 1}, \quad (4.20)$$

$$\hat{p}_{TSCORE} = \hat{q}_{TSCORE} \hat{x}^*. \quad (4.21)$$

Príklad 6: Odhad parametrov inverzného gamma rozdelenia.

Na základe definície rozdelenia a podľa (4.6) volíme $\eta = \ln x$. Potom podľa vzťahu (4.5) *t-score* pre inverzné gamma rozdelenie má tvar

$$T(x) = \alpha - \frac{\gamma}{x}.$$

Prislúchajúce charakteristiky sú

$$S(x) = \frac{\alpha}{\gamma} \left(\alpha - \frac{\gamma}{x} \right), \quad x^* = \frac{\gamma}{\alpha}, \quad \omega^2 = \frac{\gamma^2}{\alpha^3}.$$

Z (4.13) pre odhad parametrov α a γ dostaneme sústavu rovníc

$$\begin{aligned} \sum_{i=1}^n \left(\alpha - \frac{\gamma}{x_i} \right) &= 0 \\ \frac{1}{n} \sum_{i=1}^n \left(\alpha - \frac{\gamma}{x_i} \right)^2 &= \alpha. \end{aligned}$$

Dosadením x^* do prvej rovnice

$$0 = \sum_{i=1}^n \gamma \left(\frac{\alpha}{\gamma} - \frac{1}{x_i} \right) = \sum_{i=1}^n \gamma \left(\frac{1}{x^*} - \frac{1}{x_i} \right)$$

dostaneme tvar pre odhad

$$\hat{x}^* = \bar{x}_H,$$

kde $\bar{x}_H$ je harmonický priemer (4.17). Postupnými úpravami dostaneme odhady parametrov inverzného gamma rozdelenia

$$\hat{\alpha}_{TSCORE} = \frac{n}{\sum_{i=1}^n \left(1 - \frac{\bar{x}_H}{x_i} \right)^2} \quad (4.22)$$

$$\hat{\gamma}_{TSCORE} = \hat{\alpha}_{TSCORE} \bar{x}_H. \quad (4.23)$$

Kapitola 5

Aplikácia odhadov *t-score*

Na odhad parametrov rozdelení máme niekoľko metód, ktoré sme spomenuli v 3. kapitole. Cieľom tejto kapitoly je ukázať dôležitosť metódy *t-score* momentov a kvalitu jej odhadov pre rozdelenia s ťažkými chvostami. Táto metóda je pre tieto rozdelenia účinnejšia ako metóda maximálnej vierohodnosti, čo je spôsobené robustnosťou metódy *t-score* momentov.

V tejto kapitole sa pozrieme na porovnanie odhadov parametrov rozdelení získaných pomocou metódy maximálnej vierohodnosti a metódou *t-score* momentov. Ďalej sa budeme zaoberať porovnaniami odhadov v prípade odl'ahých a kontaminovaných pozorovaní pre Paretovo rozdelenie. V závere sa pozrieme na agregované nároky a ich zmenu, ak individuálne nároky budú pochádzať z rozdelenia s odhadnutými parametrami získaných pomocou týchto dvoch metód.

Simulácia dát a všetky výpočty prebiehajú v programovacom prostredí *MatLab*.

Tieto príklady s menšími úpravami, zmenenými parametrami a mnohé ďalšie môže čitateľ nájsť v [8] a [11].

5.1 Porovnanie odhadov

Porovnáme odhady získané pomocou metódy maximálnej vierohodnosti a metódou *t-score* momentov. Rozoberieme tri rozdelenia s ťažkými chvostami, a to Paretovo rozdelenie, beta rozdelenie druhého druhu a inverzné gamma rozdelenie. Použijeme analytické vyjadrenie odhadov, ktoré sme získali v kapitole 3 a kapitole 4. Odhady budeme testovať pre dané hodnoty parametrov, pričom budeme postupne zväčšovať rozsah výberu. Predovšetkým pre malé výbery bude metóda *t-score* momentov dávať v porovnaní s metódou maximálnej vierohodnosti presnejšie výsledky.

5.1.1 Paretovo rozdelenie

a) jednoduchá verzia

Ide o prípad Paretovho rozdelenia (2.4), kedy parameter $a = 1$. Tento model patrí medzi najčastejšie modely používané v poistnej matematike. Je známy pod názvom jednoparametrový Pareto model a je prebraný z práce Philbricka [10].

V predchádzajúcich kapitolách sme vypočítali odhady pre parameter c , vid' (3.2), (4.16). Pripomeňme, že

$$\hat{c}_{ML} = \frac{n}{\sum_{i=1}^n \ln x_i}, \quad \hat{c}_{TSCORE} = \frac{1}{\bar{x}_H - 1},$$

kde $\bar{x}_H$ je harmonický priemer (4.17).

Našou úlohou je vygenerovať dáta z Paretovho rozdelenia a na ich základe porovnať odhady parametrov.

Vygenerovali sme náhodný výber rozsahu N z daného rozdelenia s parametrom $c = 2$, pričom rozsah sme postupne zväčšovali. V tabuľke sú zaznamenané priemerné hodnoty odhadu po vykonaní 100 000 experimentov.

rozsah	N = 10	N = 20	N = 50	N = 100	N = 1 000	N = 10 000
odhad parametra						
$\hat{c}_{ML}$	2.2244	2.1063	2.0415	2.0209	2.0020	2.0001
$\hat{c}_{TSCORE}$	2.1705	2.0805	2.0317	2.0156	2.0015	2.0000

Tabuľka 5.1: Výsledky odhadov pre jednoduchú verziu Paretovho rozdelenia

Môžeme si všimnúť, že pre každý zvolený rozsah výberu metóda *t-score* momentov dáva presnejší odhad parametra Paretovho rozdelenia ako metóda maximálnej vierohodnosti. Už pre malý rozsah výberu je metóda *t-score* momentov efektívna. Čím menší rozsah výberu dát máme, tým je kvalita metódy *t-score* momentov v porovnaní s metódou maximálnej vierohodnosti väčšia. So zväčšujúcim rozsahom výberu dávajú obe metódy presnejšie odhady a pri výbere 10 000 sú ich odhady takmer presné.

Keďže jednoparametrové rozdelenie patrí medzi najpoužívanejšie rozdelenie s ťažkými chvostami v poistnej matematike pri modelovaní očakávaných nárokov, dôležitosť a efektívnosť metódy *t-score* momentov sa nám podarilo dokázať.

b) zovšeobecnená verzia

Pripomeňme tvary odhadov parametrov a a c získané metódou maximálnej vierohodnosti, vid' (3.2), (3.3)

$$\hat{a}_{ML} = \min_i x_i, \quad \hat{c}_{ML} = \frac{n}{\sum_{i=1}^n (\ln x_i - \ln \hat{a}_{ML})}$$

a získané metódou *t-score* momentov, vid' (4.18), (4.19)

$$\hat{c}_{TSCORE} = -1 + \sqrt{1 + \frac{1}{A}}, \quad \hat{a}_{TSCORE} = \frac{\hat{c}_{TSCORE}}{\hat{c}_{TSCORE} + 1} \bar{x}_H,$$

kde $\bar{x}_H$ je harmonický priemer (4.17) a $A = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\bar{x}_H}{x_i}\right)^2$.

Našou úlohou je opäť vygenerovať dáta z Paretovho rozdelenia a na ich základe porovnať odhady parametrov.

Generovali sme výber z daného rozdelenia s parametrami $a = 0.5$ a $c = 2$. Postup bol rovnaký ako v predošlom prípade a výsledky sú zaznamenané v nasledujúcej tabuľke.

rozsah	N = 10	N = 20	N = 50	N = 100	N = 1 000	N = 10 000
odhad parametrov						
$\hat{a}_{ML}$	0.5263	0.5128	0.5050	0.5025	0.5002	0.5000
$\hat{c}_{ML}$	2.5012	2.2212	2.0845	2.0399	2.0039	2.0004
$\hat{a}_{TSCORE}$	0.5155	0.5072	0.5027	0.5014	0.5001	0.5000
$\hat{c}_{TSCORE}$	2.4178	2.1765	2.0661	2.0309	2.0030	2.0004

Tabuľka 5.2: Výsledky odhadov pre zovšeobecnenú verziu Paretovho rozdelenia

Z hľadiska kvality sme dostali rovnaké výsledky ako v predchádzajúcej časti. Z tabuľky 5.2 je vidieť, že metóda *t-score* momentov je presnejšia ako metóda maximálnej vierohodnosti. So zväčšujúcim rozsahom výberu dávajú obe metódy presnejšie odhady, pričom parameter a dosahuje lepšie odhady ako parameter c . Opäť môžeme poznamenať, že pre malé výbery je rozdiel v efektívnosti metód viditeľnejší, metóda *t-score* momentov dáva oveľa lepšie výsledky.

Na základe výsledkov môžeme vyhlásiť, že metóda *t-score* momentov je vhodnejšou metódou na odhad parametrov Paretovho rozdelenia ako metóda maximálnej vierohodnosti.

5.1.2 Inverzné gamma rozdelenie

Máme analytické vyjadrenie odhadov parametrov α a γ získané metódou t -score momentov, vid' (4.22), (4.23)

$$\hat{\alpha}_{TSCORE} = \frac{n}{\sum_{i=1}^n \left(1 - \frac{\bar{x}_H}{x_i}\right)^2}, \quad \hat{\gamma}_{TSCORE} = \hat{\alpha}_{TSCORE} \bar{x}_H,$$

kde $\bar{x}_H$ je harmonický priemer (4.17).

Odhad parametrov pomocou metódy maximálnej vierohodnosti získame transformáciou dát a použitím funkcie *gamfit* pre gamma rozdelenie, ktorá je súčasťou *MatLab-u*.

Generovali sme výber z daného rozdelenia s parametrami $\alpha = 2.5$ a $\gamma = 4$, pričom rozsah výberu sme postupne zväčšovali. Priemerné hodnoty odhadov po vykonaní 10 000 experimentov sú zaznamenané v nasledujúcej tabuľke. Počet simulácií sme znížili pre časovú náročnosť príkladu.

rozsah	N = 10	N = 20	N = 50	N = 100	N = 1 000	N = 10 000
odhad parametrov						
$\hat{\alpha}_{ML}$	3.4630	2.9257	2.6491	2.5737	2.5067	2.5004
$\hat{\gamma}_{ML}$	4.4401	4.2402	4.0790	4.0507	4.0032	4.0003
$\hat{\alpha}_{TSCORE}$	3.7064	3.0600	2.7144	2.6070	2.5108	2.5008
$\hat{\gamma}_{TSCORE}$	6.1798	4.9976	4.3713	4.1920	4.0185	4.0019

Tabuľka 5.3: Výsledky odhadov pre inverzné gamma rozdelenie

Metóda t -score momentov dosiahla v tomto prípade horšie výsledky odhadov parametrov ako metóda maximálnej vierohodnosti. Metóda maximálnej vierohodnosti je vhodnejšia na odhad parametrov inverzného gamma rozdelenia.

5.1.3 Beta rozdelenie druhého druhu

Máme odhady parametrov p a q získané metódou t -score momentov, vid' (4.21), (4.20)

$$\hat{q} = \frac{A\hat{x}^* - 1}{\hat{x}^* + 1}, \quad \hat{p} = \hat{q} \hat{x}^*,$$

$$\text{kde substitúcie sú rovné } \hat{x}^* = \frac{\sum_{i=1}^n \frac{x_i}{x_i+1}}{\sum_{i=1}^n \frac{1}{x_i+1}} \quad \text{a} \quad A = \frac{n}{\sum_{i=1}^n \left(\frac{x_i - \hat{x}^*}{x_i+1}\right)^2}.$$

Odhad parametrov pomocou metódy maximálnej vierohodnosti získame transformáciou dát a

použitím funkcie *betafit* pre beta rozdelenie, ktorá je súčasťou *MatLab-u*.

Generovali sme výber z daného rozdelenia s parametrami $p = 2.5$ a $q = 5$, pričom rozsah výberu sme postupne zväčšovali. Priemerné hodnoty odhadov po vykonaní 10 000 experimentov sú zaznamenané v nasledujúcej tabuľke. Počet simulácií sme znížili pre časovú náročnosť príkladu.

rozsah	N = 10	N = 20	N = 50	N = 100	N = 1 000	N = 10 000
odhad parametrov						
$\hat{\alpha}_{ML}$	3.4780	2.9179	2.6513	2.5746	2.5056	2.5008
$\hat{\lambda}_{ML}$	7.0909	5.8857	5.3249	5.1629	5.0134	5.0022
$\hat{\alpha}_{TSCORE}$	3.4479	2.9041	2.6493	2.5727	2.5055	2.5007
$\hat{\gamma}_{TSCORE}$	7.0253	5.8525	5.3173	5.1576	5.0130	5.0020

Tabuľka 5.4: Výsledky odhadov pre beta rozdelenie druhého druhu

Z tabuľky 5.4 je vidieť, že obe metódy dávajú pre určitý rozsah výberu približne rovnaké výsledky, avšak opäť metóda *t-score* momentov je pre odhad parametrov beta rozdelenia druhého druhu efektívnejšia.

V tejto časti sme ukázali, že metóda *t-score* momentov na odhad parametrov rozdelení je pre rozdelenia s ťažkými chvostami efektívnejšia ako metóda maximálnej vierohodnosti. Potvrdilo sa to najmä pre Paretovo rozdelenie. Pre beta rozdelenie druhého druhu je táto metóda tiež efektívnejšia, avšak rozdiely sú už menšie. Pre inverzné gamma rozdelenie je metóda odhadu pomocou maximálnej vierohodnosti lepšia, rozdiely sa ukázali hlavne pri malom rozsahu výberu.

5.2 Odľahlé pozorovania

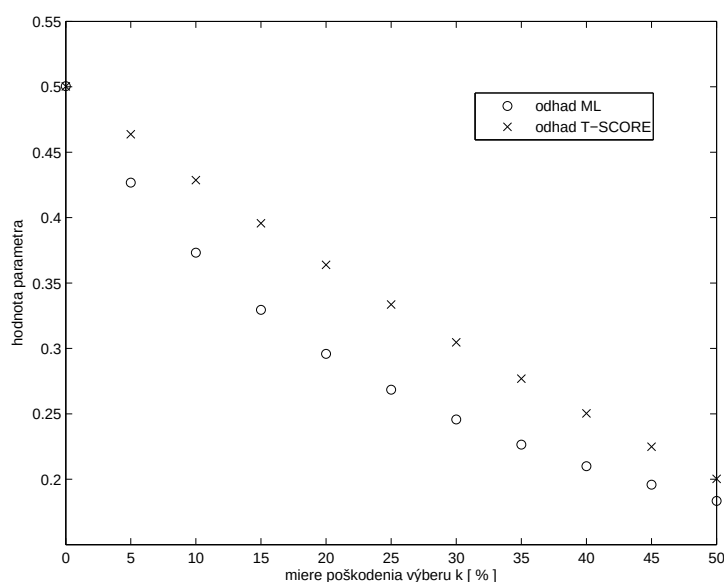
V tejto časti sa budeme zaoberať porovnaniami odhadov získanými metódou maximálnej vierohodnosti a metódou *t-score* momentov za prítomnosti outlierov medzi pozorovaniami. Odhady budeme testovať na jednoparametrovom Paretovom rozdelení, kde odhady parametrov majú tvar (3.2) a (4.16).

Vygenerovali sme náhodný výber rozsahu 1 000 z daného rozdelenia s parametrom c . Odľahlé pozorovania sme dostali odobratím $k\%$ dát a vynásobením 1 000, pričom za k sme zobrali 10% a 30%. Parameter c sme zvolili $c = 0.5$ a neskôr $c = 2.5$. Pre každý prípad sme urobili 100 000 experimentov a priemerné výsledky sú zaznamenané v nasledujúcej tabuľke.

parameter	c = 0.5		c = 2.5	
množstvo outlierov	k = 10%	k = 30%	k = 10%	k = 30%
$\hat{c}_{ML}$	0.3717	0.2457	0.9169	0.4045
$\hat{c}_{TSCORE}$	0.4286	0.3046	1.8015	1.0013

Tabuľka 5.5: Výsledky odhadov pre odľahlé pozorovania

Z tabuľky 5.5 je vidieť, že metóda *t-score* momentov dáva presnejšie odhady ako metóda maximálnej vierohodnosti pre ľubovoľnú hodnotu parametra *c* aj ľubovoľnú mieru poškodenia výberu *k*.



Obr. 5.1: Graf odhadov parametra Paretovho rozdelenia pre rôzne miery poškodenia výberu odľahlými pozorovaniami

Z predchádzajúceho grafu 5.1 vyplýva, že ak sa zväčšuje počet odľahlých pozorovaní, tak rozdiel medzi odhadnutými parametrami sa zväčšuje, teda metóda *t-score* momentov je efektívnejšia ako metóda maximálnej vierohodnosti. Pri 30% poškodení sa rozdiely začnú zmenšovať.

5.3 Kontaminované pozorovania

V tejto časti sa budeme zaoberať porovnaniami odhadov získanými metódou maximálnej vierohodnosti a metódou *t-score* momentov za prítomnosti kontaminovaných dát medzi

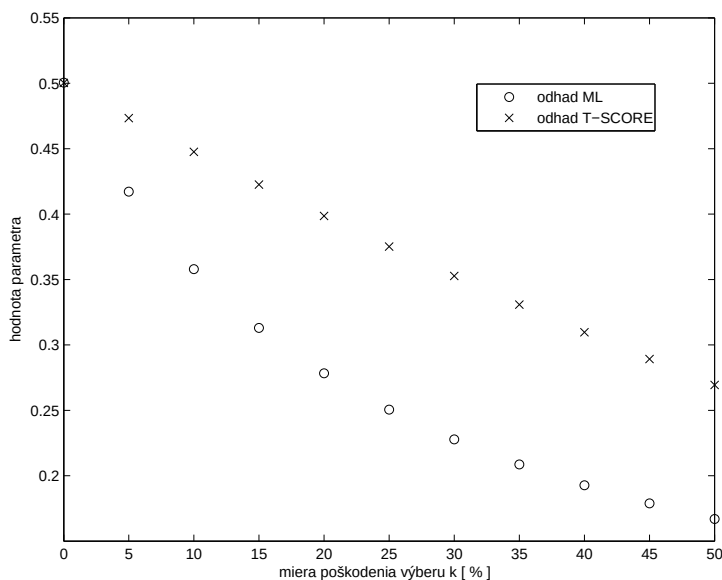
pozorovaniami. Odhady budeme testovať na jednoparametrovom Paretovom rozdelení, kde odhady parametrov majú tvar (3.2) a (4.16).

Vygenerovali sme náhodný výber rozsahu 1 000, pričom kontaminované dáta sú z Paretovho rozdelenia s parametrom $c = 0.1$ a tvoria $k\%$ z náhodného výberu a zvyšok je z Paretovho rozdelenia s parametrom c . Za k sme zobrali 10% a 30%. Parameter c sme zvolili $c = 0.5$ a neskôr $c = 2.5$. Pre každý prípad sme urobili 100 000 experimentov a priemerné výsledky sú zaznamenané v nasledujúcej tabuľke.

parameter	$c = 0.5$		$c = 2.5$	
množstvo kontaminovaných dát	$k = 10\%$	$k = 30\%$	$k = 10\%$	$k = 30\%$
$\hat{c}_{ML}$	0.3579	0.2278	0.7397	0.3058
$\hat{c}_{TSCORE}$	0.4476	0.3527	1.8745	1.1160

Tabuľka 5.6: Výsledky odhadov pre kontaminované pozorovania

Z tabuľky 5.6 je vidieť, že metóda t -score momentov dáva presnejšie odhady ako metóda maximálnej vierohodnosti pre ľubovoľnú hodnotu parametra c aj ľubovoľnú mieru poškodenia výberu k .



Obr. 5.2: Graf odhadov parametra Paretovho rozdelenia pre rôzne miery poškodenia výberu kontaminovanými pozorovaniami

Z predchádzajúceho obrázka 5.2 vyplýva, že ak sa zväčšuje počet kontaminovaných pozorovaní, tak rozdiel medzi odhadnutými parametrami sa zväčšuje. Metóda *t-score* momentov je efektívnejšia ako metóda maximálnej virohodnosti pri zväčšujúcej sa miere poškodenia výberu v prípade kontaminovaných dát.

5.4 Agregované nároky

V tejto časti sa budeme zaoberať správaním agregovaných nárokov. V praxi sa ukazuje, že rozdelenie sumy agregovaných nárokov závisí od rozdelenia jednotlivých nárokov, preto je dôležité používať robustné odhady.

Uvažujeme model

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

kde S je súčet N nárokov X_i . Nároky X_i pochádzajú zo spojitých rozdelení definovaných v 2. kapitole. Počet nárokov N je konštantný (to sa však v praxi nepredpokladá) alebo náhodný, pričom jeho rozdelenie patrí do skupiny diskretných rozdelení, napr. binomické rozdelenie, negatívne binomické rozdelenie alebo Poissonovo rozdelenie.

V tomto príklade rozoberieme sumu individuálnych nárokov, ktoré pochádzajú z loggamma rozdelenia. Poukážeme na to, aký vplyv na rozdelenie sumy nárokov majú odhady parametrov loggamma rozdelenia získaných pomocou metódy maximálnej virohodností a metódou *t-score* momentov.

Vygenerovali sme náhodný výber rozsahu $N = 100$ z loggamma rozdelenia s parametrami $\alpha = 10$ a $\gamma = 2.5$. Spravili sme 100 000 simulácií. Z týchto dát sme odhadli parametre rozdelenia, pričom sme použili odhady získané metódou *t-score* momentov, viď (4.14), (4.15)

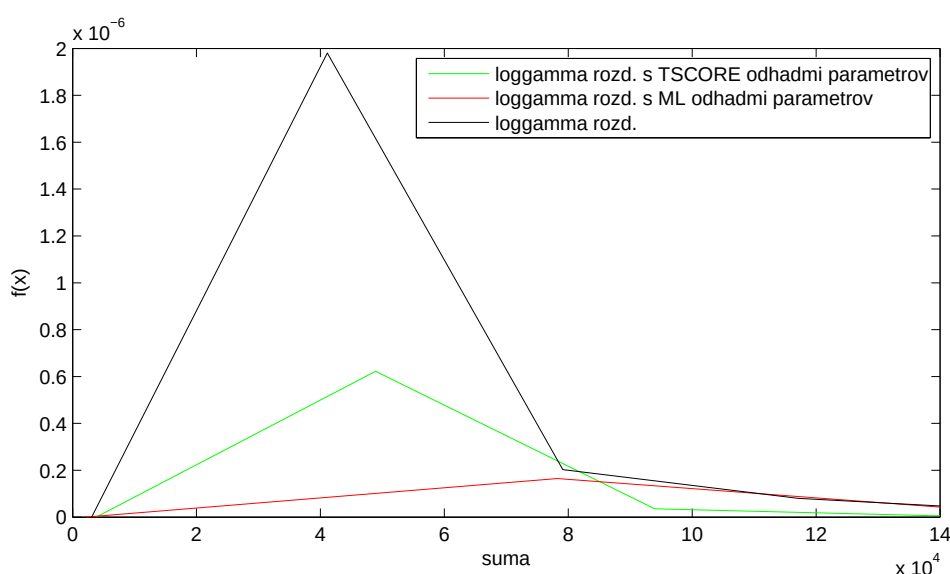
$$\hat{c}_{TSCORE} = \frac{\sum_{i=1}^n \ln x_i}{\sum_{i=1}^n (\ln x_i)^2 - \frac{1}{n} \left(\sum_{i=1}^n \ln x_i \right)^2}, \quad \hat{\alpha}_{TSCORE} = \hat{c}_{TSCORE} \frac{\sum_{i=1}^n \ln x_i}{n}.$$

Pri odhade parametrov $\hat{\alpha}_{ML}$ a $\hat{\gamma}_{ML}$ metódou maximálnej virohodnosti použijeme transformáciu dát a na ne aplikujeme *Matlab-ovskú* funkciu *gamfit*. Dostali sme štyri rôzne výsledky pre sumu 100 nárokov pochádzajúcich z loggamma rozdelenia.

1. Teoretický výpočet : $Suma = 100(1 - \frac{1}{\gamma})^{-\alpha} = 16\,538.17$.
2. $Suma = 16\,570.09$ pre nároky generované z loggamma rozdelenia s parametrami α a γ .

3. $Suma = 17\,438.63$ pre nároky generované z loggamma rozdelenia s parametrami $\hat{\alpha}_{ML}$ a $\hat{\gamma}_{ML}$.
4. $Suma = 17\,353.621$ pre nároky generované z loggamma rozdelenia s parametrami $\hat{\alpha}_{TSCORE}$ a $\hat{\gamma}_{TSCORE}$.

Z výsledkov je vidieť, že metóda odhadu t -score momentov dáva lepšie výsledky ako metóda maximálnej vierohodnosti. Suma nárokov z loggamma rozdelenia s parametrami odhadnutými pomocou metódy t -score je bližšie k teoretickej hodnote ako suma nárokov pochádzajúcich z loggamma rozdelenia s parametrami odhadnutými metódou maximálnej vierohodnosti.



Obr. 5.3: Graf hustoty sumy individuálnych nárokov

V grafe 5.3 je zobrazená hustota sumy nárokov pochádzajúcich z loggamma rozdelenia s danými parametrami α a γ a parametrami získanými metódou maximálnej vierohodnosti a metódou t -score momentov. Aj graficky je potvrdené, že metóda t -score momentov je v prípade agregovaných nárokov efektívnejšia. Hustota sumy nárokov pochádzajúcich z loggamma rozdelenia s parametrami odhadnutými pomocou t -score momentov sa menej odlišuje od hustoty sumy nárokov pochádzajúcich z loggamma rozdelenia s parametrami α a γ ako suma nárokov pochádzajúcich z loggamma rozdelenia s parametrami odhadnutými pomocou maximálnej vierohodnosti.

Záver

Témou tejto diplomovej práce boli štatistické odhady parametrov rozdelení v poisťovníctve. Spomínaná problematika je dôležitá pre poisťovne pri modelovaní očakávaných nárokov a následne pri správnom určení výšky poistného.

V práci sme spracovali základné poznatky o najznámejších metódach odhadov parametrov rozdelení, a to o metóde maximálnej vierohodnosti, metóde najmenších štvorcov a metóde momentov. Najväčšiu pozornosť sme venovali metóde *t-score* momentov.

Cieľom práce bolo ukázať efektívnosť metódy *t-score* momentov v prípade rozdelení s ťažkými chvostami. Táto trieda rozdelení je vhodná na popis vysokých poistných nárokov, s ktorými sa poisťovne v poslednom čase stretávajú v dôsledku prírodných nešťastí a veľkých nehôd.

Porovnali sme odhady získané metódou maximálnej vierohodnosti a metódou *t-score* momentov pre Paretovo rozdelenie, beta rozdelenie druhého druhu a inverzné gamma rozdelenie. V prípade Paretovho rozdelenia zaznamenala metóda *t-score* momentov lepšie výsledky, už pre malý rozsah výberu boli výsledky uspokojivé. Aj pre beta rozdelenie druhého druhu sa táto metóda ukázala efektívnejšia. V prípade inverzného gamma rozdelenia bola efektívnejšia metóda maximálnej vierohodnosti.

Ďalej sme sa venovali analýze senzitivnosti pre Paretovo rozdelenie, ktoré patrí medzi najpoužívanéjšie rozdelenie v neživotnom poisťovníctve. Zvolili sme špeciálny typ tohto rozdelenia, jednoparametrové Paretovo rozdelenie. Rozobrali sme dva prípady. V prvom sme sa venovali výberu s odľahlými pozorovaniami a v druhom výberu s kontaminovanými pozorovaniami. V oboch prípadoch dosiahla metóda *t-score* lepšie výsledky. V prípade poškodenia výberu kontaminovanými dátami dosahovala táto metóda v porovnaní s metódou maximálnej vierohodnosti väčšiu efektívnosť s rastúcou mierou poškodenia výberu.

V poslednej časti sme sa pozreli, ako tieto dve metódy vplývajú na agregované nároky. Podarilo sa nám ukázať, že suma nárokov pochádzajúcich z loggamma rozdelenia s parametrami odhadnutými pomocou metódy *t-score* bola bližšie k teoretickej hodnote ako suma nárokov pochádzajúcich z loggamma rozdelenia s parametrami odhadnutými metódou maximálnej vierohodnosti.

Analytické tvary *t-score* odhadov, ktoré sme použili pri skúmaní, sme vyjadrili v štvrtej kapitole. Odhady maximálnej vierohodnosti sme dostali transformáciu dát a použitím *MatLab-ovských* funkcií *betafit* a *gamfit*, pre Paretovo rozdelenie sa nám podarilo vyjadriť analytické tvary aj pomocou tejto metódy. Okrem týchto odhadov boli v práci spracované aj odhady získané metódou najmenších štvorcov pre Makehamovo rozdelenie.

Na základe našich dosiahnutých výsledkov môžeme zhodnotiť, že sa nám podarilo ukázať efektivitu metódy *t-score* momentov pre rozdelenia s ťažkými chvostami. V porovnaní s metódou maximálnej vierohodnosti dosiahla lepšie výsledky.

Literatúra

- [1] EMBRECHTS, P., KLÜPPELBERG, C., MIKOSCH, T. *Modelling extremal events for insurance and finance*. 648s. Berlín, Springer-Verlang, 2003. ISBN 3-540-60931-8.
- [2] FABIÁN, Z. *Estimation of simple characteristic of samples skewed and heavy-tailed distributions*. 8s. Institute of Computer Sciency. 2006.
- [3] FABIÁN, Z. *Johnson score and characteristics of distributions*. 10s. Institute of Computer Sciency, Academy of Sciences of the Czech republic, No.963. 2006.
- [4] FABIÁN, Z. *Johnson point and Johnson variance*. 10s. Institute of Computer Sciency, Academy of Sciences of the Czech republic, No.963. 2006.
- [5] FABIÁN, Z. *Parametric estimation by generalized moment method for etremes*. 12s. Institute of Computer Sciency. 2006.
- [6] FABIÁN, Z. *New measures of central tendency and variability of continiuous distributions*. Communications in Statistics - Theory and Methods, Volume 37, Issue 2, 159s.-174s. 2008.
- [7] FENG, X., HE, G., ABDURISHIT. *Estimation of parameters of the Makeham distribution using the least squares method*. 34s. - 44s. 2006.
- [8] KEPLER, J., FABIÁN, Z., STEHLÍK, M. *A note on favorable estimation when data is contaminated*. 8s. 2008.
- [9] LAMOŠ, F., POTOCKÝ, R. *Pravdepodobnost' a matematická štatistika*. Bratislava. Vydavateľ'stvo UK. 344s. 1998. ISBN 80-223-1262-2.
- [10] PHILBRICK, S. *A practical guide to the single parameter Pareto distribution*. 1985.
- [11] STEHLÍK, M., POTOCKÝ, R., WALDL, H., FABIÁN Z. *Some notes on the favorable estimation for fitting heavy tailed data*. IFAS Research Paper Series 2008 - 32 . 30s. 2008.

- [12] KOLLÁR, M., KOSSACZKÁ, L., ŠEVČOVIČ, D. *Diferenciálny a integrálny počet viac premenných v príkladoch*. Bratislava. Knížničné a edičné centrum FMFI UK. 2010. 192s. ISBN 978-80-89186-54-9.
- [13] VIDMAR, G., BLAGUS, R., STŘELEČ, L, STEHLÍK, M. *Business indicators of healthcare quality: Outlier detection in small samples*. 14s. Applied Stochastic Models in Business and Industry. 2011.