VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistro darbas

Kai kurios sudėtinio lognormaliojo – apibendrinto Pareto skirstinio savybės

Some properties of a composite lognormal – generalized Pareto distribution

Gediminas Kuodis

VILNIUS 2008

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATINĖS ANALIZĖS KATEDRA

Darbo vadovas *Lietuvos MA narys korespondentas Habil. dr.*

Vygantas Paulauskas _______________
(parašas)

Darbas apgintas ____2008 m. gegužės mėn. 29 d.___
Gynimo posėdžio protokolo Nr. _________________
Darbas įvertintas ______________________________

Registravimo Nr. ________________
2008-05-25 ____________________

Turinys

ANOTACIJA/REZIUMĖ

Šis darbas remiasi dviem straipsniais: Kahadawala Cooray, Malwane M. A. Amanda, „Modeling actuarial data with a composite lognormal – Pareto model“ (Scandinavian Actuarial Journal, 5, 321-334 psl.) ir McNeil Alexander J. „Estimating the tails of loss severity distributions using extreme value theory“ (ASTIN Bulletin, 27, 117-137 psl.). Pirmajame, kuris buvo nagrinėtas bakalauro darbe, pristatomas sudėtinis lognormalusis – Pareto skirstinys. Antrajame nagrinėjamas apibendrintas Pareto skirstinys, tiriama, kaip jis aprašo dideles žalas.

Šio magistro darbo tikslas yra sujungti lognormalųjį bei apibendrintą Pareto skirstinius. Pirmasis jų gerai aprašo mažas žalas su dideliais dažniais, antrasis – dideles, turinčias mažus dažnius. Darbe taip pat naudojamas kiek pakeistas Kahadawala Cooray ir Malwane M. A. Amanda straipsnyje pasiūlytas skirstinių sujungimo būdas. Šiame darbe ištirtos kai kurios sudėtinio lognormaliojo – apibendrinto Pareto skirstinio su keturiais laisvais parametrais savybės, pateiktas įverčių radimo metodas, bei, remiantis šiuo modeliu, išnagrinėtos trys duomenų imtys.

Rezultatai rodo jog, dėl tam tikrų sujungimo savybių, modelis tinkamas aprašyti tiek didelėms, tiek mažoms žaloms, o, svarbiausia, mišrioms žaloms, tarp kurių pasitaiko tiek mažų, tiek labai didelių. Pastarojo tipo duomenys gana dažni draudimo praktikoje.

This work is based on two articles: Kahadawala Cooray, Malwane M. A. Amanda, „Modeling actuarial data with a composite lognormal – Pareto model“ (Scandinavian Actuarial Journal, 5, pages 321-334) and McNeil Alexander J. „Estimating the tails of loss severity distributions using extreme value theory“ (ASTIN Bulletin, 27, pages 117-137). The first article presents a two - parameter smooth continuous composite lognormal - Pareto model. The second one analyses generalised Pareto distribution and how does this distribution cover large loses.

The purpose of this writing is to mix lognormal and generalised Pareto distributions in to one four parameter smooth continuous distribution. The lognormal distribution is used to model small data with higher frequencies, and the generalised Pareto distribution is used to model large data with low frequencies. In this work we also try to use a modified way of mixing these two distributions than Kahadawala Cooray and Malwane M. A. Amanda used in their work. Writing also includes an analysis of some properties of this model, method of parameter estimation and three data sets analysis based on this distribution.

The results show that because of some properties of mixing these two distributions, a composite model is suitable for covering small loses as well as it is suitable for covering large loses. But most important result is that the composite lognormal - generalised Pareto distribution is fit to cover data sets, which include small loses and very large loses also. This type of data is very common in an insurance industry.

ĮVADAS

Šiame magistro darbe nagrinėjama gana svarbi draudimo rinkose iškylanti problema: dažniausiai draudimo kompanijų žalas sudaro daug nelabai didelių žalų, kurias gerai aprašo lognormalusis skirstinys, tačiau tikėtinos ir labai didelės žalos, kurių dažnis nėra toks aukštas, tačiau to užtenka, kad lognormalusis skirstinys nesugebėtų pakankamai įvertinti galimų draudimo nuostolių. Didelėms žaloms tinkamesni yra Pareto arba apibendrintas Pareto skirstiniai, kurie turi gana sunkias uodegas. Jie tinkami aprašyti dideles žalas ir yra ypač aktualūs perdraudimo srityje. Tačiau pastarieji du skirstiniai netinka vertinti mažoms žaloms. Taigi, siekiant aprašyti visas žalas (tiek mažas, su dideliais dažniais, tiek dideles, su mažais dažniais), reikia rasti skirstinį, turintį longnormaliajam tankiui artimą formą mažoms žaloms vertinti, tačiau jo uodega turi būti Pareto tipo (tai yra sunki), kad sugebėtų įvertinti dideles žalas, pasirodančias santykinai retai.

Bakalauro darbe mes nagrinėjome sudėtinį lognormalųjį – Pareto skirstinį, turintį du laisvus parametrus, aprašytą Kahadawala Cooray ir Malwane M. A. Amanda straipsnyje „Modeling actuarial data with a composite lognormal – Pareto model“. Šiame magistro darbe sukurtas sudėtinis lognormalusis – apibendrintas Pareto modelis. Pareto skirstinio pakeitimas apibendrintu Pareto leidžia padidinti laisvų parametrų skaičių nuo dviejų iki trijų. Taip pat pakeistas sujungimo būdas: įvestas papildomas svorio parametras d ($0 \leq d \leq 1$), priskiriantis lognormaliajam skirstiniui svorį tarp 0 ir 1, atitinkamai, apibendrintam Pareto skirstiniui priskiriant svorį $1\text{-}d$. Toks sujungimas, leidžiant parametrui kisti tarp 0 ir 1, suteikia skirstiniui daugiau lankstumo, ieškant konkrečius duomenis atitinkančių įverčių. Šis tolydus sudėtinis lognormalusis – apibendrintas Pareto skirstinys turi keturis laisvus parametrus $\alpha > 0$, $\beta > 0$, $\theta > 0$, $\sigma > 0$. Iki parametro θ jis įgyja lognormaliojo tankio formą, o į dešinę nuo jo - apibendrinto Pareto tankio formą.

Pirmame darbo skyriuje detaliai pateikta, kaip išvedamas šis skirstinys. Parodyta, kaip suskaičiuojamas jo pasiskirstymas, momentai. Antrame skyriuje pateiktas įverčių radimo algoritmas, bei funkcija, skirta skaičiuoti šio skirstinio įverčiams statistiniu paketu „R“. Trečiame skyriuje, naudojantis Kolmogorovo – Smirnovo, Anderseno – Darlingo testų statistikomis, bei Chi kvadrato suderinamumo kriterijaus p – reikšme, įvertinta, kaip šis modelis atitinka realius duomenis. Išnagrinėtos trys imtys: sudėtiniu lognormaliuoju – Pareto modeliu sugeneruoti duomenys, Danijos gaisrų draudimo žalos, bei vienos Lietuvos draudimo kompanijos ne gyvybės draudimo žalos.

1. SUDĖTINIS LOGNORMALUSIS – APIBENDRINTAS PARETO MODELIS

1.1. Tankio funkcija

Sakykime, *X* yra atsitiktinis dydis turintis tankio funkciją:

$$f(x)=\begin{cases}\dfrac{df_1(x)}{F_1(\theta)}, & \text{jei } 0<x\le\theta,\\ (1-d)\,f_2(x), & \text{jei } \theta\le x<\infty,\end{cases}$$

čia $\dfrac{f_1(x)}{F_1(\theta)}$ - taške θ nupjautas lognormalusis tankis su dviem parametrais, $f_2(x)$ - apibendrintas Pareto tankis su trimis parametrais, t.y.

$$\frac{f_1(x)}{F_1(\theta)}=\frac{1}{F_1(\theta)}\frac{(2\pi)^{-1/2}}{x\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\},\ x>0,$$

$$f_2(x)=\frac{1}{\beta}\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha-1},\ x>\theta,$$

kur μ, σ, α, β, θ yra nežinomi parametrai ir $\sigma>0$, $\alpha>0$, $\beta>0$, $\theta>0$.

Kadangi lognormalusis skirstinys apibrėžiamas intervale nuo 0 iki ∞, o šiuo atveju jis imamas tik iki taško θ, jis nupjaunamas šiame taške, t.y. tankį daliname iš pasiskirstymo funkcijos šiame taške. Lognormaliojo skirstinio pasiskirstymo funkciją taške θ galime išreikšti per normaliojo skirstinio pasiskirstymo funkciją tokiu būdu:

$$I:=\int_0^{\theta} f_1(x)\,dx=\int_0^{\theta}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\}dx;$$

Atliekame pakeitimą: $\mathrm{t}=\dfrac{\ln x-\mu}{\sigma}$;

$$I=\int_{-\infty}^{\frac{\ln\theta-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}dx=\Phi\left(\frac{\ln\theta-\mu}{\sigma}\right).$$

$\Phi(\bullet)$ yra standartinio normaliojo atsitiktinio dydžio pasiskirstymo funkcija. Tokiu būdu gauname nupjautą lognormalųjį skirstinį, apibrėžiamą intervale nuo 0 iki θ. Apibendrintas Pareto skirstinys apibrėžiamas nuo taško θ iki ∞, taigi, jo keisti mums nereikia.

Parametras d reguliuoja, koks svoris tenka kiekvienam iš skirstinių, jis apibrėžiamas intervale [0;1]. Kadangi atskirai kiekvienas iš skirstinių pilnai nusako atsitiktinius dydžius atitinkamai nuo 0 iki θ ir nuo θ iki ∞, tai dauginant jų tankio funkcijas atitinkamai iš d ir 1-d, gauname sudėtinį skirstinį, nusakantį atsitiktinį dydį visame intervale (0, ∞).

Pareikalausime, kad funkcijos $f_1(x)$ ir $f_2(x)$ būtų tolydžios ir diferencijuojamos taške θ. Sudarome lygčių sistemą:

$$\begin{cases} \dfrac{df_1(\theta)}{\Phi\left(\dfrac{\ln\theta-\mu}{\sigma}\right)} = (1-d)\, f_2(\theta), \\ \dfrac{df'_1(\theta)}{\Phi\left(\dfrac{\ln\theta-\mu}{\sigma}\right)} = (1-d)\, f'_2(\theta), \end{cases}$$

kur $f'(\theta)$ yra pirmoji $f(x)$ išvestinė taške θ. Šios dvi sąlygos garantuoja tolydžią ir tolygiai diferencijuojamą funkciją. Apribojimai taške θ sumažina nežinomų parametrų skaičių nuo šešių iki keturių.

Raskime funkcijų $f_1(x)$ ir $f_2(x)$ išvestines pagal x:

$$f'_1(x) = \left(\frac{1}{\sqrt{2\pi}x\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\}\right)'_x =$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\left(\left(x^{-1}\right)'\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\} + x^{-1}\left(\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\}\right)'\right) =$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\left(-\frac{1}{x^2}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\} + \frac{1}{x}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\}\left(-\frac{1}{2}\right)2\left(\frac{\ln x-\mu}{\sigma}\right)\frac{1}{x}\frac{1}{\sigma}\right) =$$

$$= -\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\}\left(1+\frac{\ln x-\mu}{\sigma^2}\right);$$

$$f'_2(x) = \left(\frac{1}{\beta}\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha-1}\right)'_x = -\frac{\alpha+1}{\beta}\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha-2}\frac{1}{\alpha\beta} = -\frac{\alpha+1}{\alpha\beta^2}\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha-2}.$$

Įsistatę funkcijas $f_1(x)$ ir $f_2(x)$ bei jų išvestines pagal x taške θ, į lygčių sistemą gauname:

$$\begin{cases} \dfrac{d\dfrac{1}{\sqrt{2\pi}\theta\sigma}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\ln\theta-\mu}{\sigma}\right)^2\right\}}{\Phi\left(\dfrac{\ln\theta-\mu}{\sigma}\right)}=(1-d)\dfrac{1}{\beta}, \\ \dfrac{d\left(-\dfrac{1}{\sqrt{2\pi}\theta^2\sigma}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\ln\theta-\mu}{\sigma}\right)^2\right\}\left(1+\dfrac{\ln\theta-\mu}{\sigma^2}\right)\right)}{\Phi\left(\dfrac{\ln\theta-\mu}{\sigma}\right)}=(1-d)\left(-\dfrac{\alpha+1}{\alpha\beta^2}\right), \end{cases}$$

Spręsdami šią sistemą turime:

$$(1-d)\frac{1}{\beta}\left(1+\frac{\ln\theta-\mu}{\sigma^2}\right)\frac{1}{\theta}=(1-d)\left(\frac{\alpha+1}{\alpha\beta^2}\right);$$

$$\sigma^2+\ln\theta-\mu=\frac{(\alpha+1)\theta\sigma^2}{\alpha\beta};$$

$$\mu=\sigma^2+\ln\theta-\frac{(\alpha+1)\theta\sigma^2}{\alpha\beta};$$

Gautą parametro μ išraišką statome į pirmąją lygčių sistemos lygtį:

$$\frac{\dfrac{d}{\sqrt{2\pi}\theta\sigma}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}-\sigma^2}{\sigma}\right)^2\right\}}{\Phi\left(\dfrac{\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}-\sigma^2}{\sigma}\right)}=(1-d)\frac{1}{\beta};$$

$$d\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}=(1-d)\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma;$$

$$d\left(\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma\right)=\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma;$$

Galiausiai gauname tokią parametro d išraišką:

$$d=\frac{\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma};$$

Taigi toks svoris priskiriamas lognormaliojo skirstinio tankiui. Apibendrinto Pareto skirstinio tankiui priskiriamas svoris 1-d :

$$1-d=\frac{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma};$$

Parametrus d ir μ išsireiškėme per likusius keturis parametrus α, β, θ, σ. Dabar galime įsistayti šias išraiškas į pradinę sudėtinio lognormaliojo - apibendrinto Pareto tankio išraišką:

$$f(x)=\begin{cases}\dfrac{\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\times \\ \times\dfrac{\frac{1}{\sqrt{2\pi}x\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\ln\theta-\sigma^2+\frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}}{\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)}, & \text{jei } 0<x\le\theta, \\ \dfrac{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\dfrac{1}{\beta}\left(1+\dfrac{x-\theta}{\alpha\beta}\right)^{-\alpha-1}, & \text{jei } \theta\le x<\infty,\end{cases}$$

Atlikę kai kuriuos prastinimo veiksmus gauname galutinę sudėtinio lognormaliojo - apibendrinto Pareto skirstinio tankio išraišką:

$$
f(x)=\begin{cases}\dfrac{\dfrac{\theta}{x}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}}{\beta\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}, & \text{jei } 0<x\le\theta,\\[2ex] \dfrac{\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}\left(1+\dfrac{x-\theta}{\alpha\beta}\right)^{-\alpha-1}}{\beta\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}, & \text{jei } \theta\le x<\infty,\end{cases}
$$

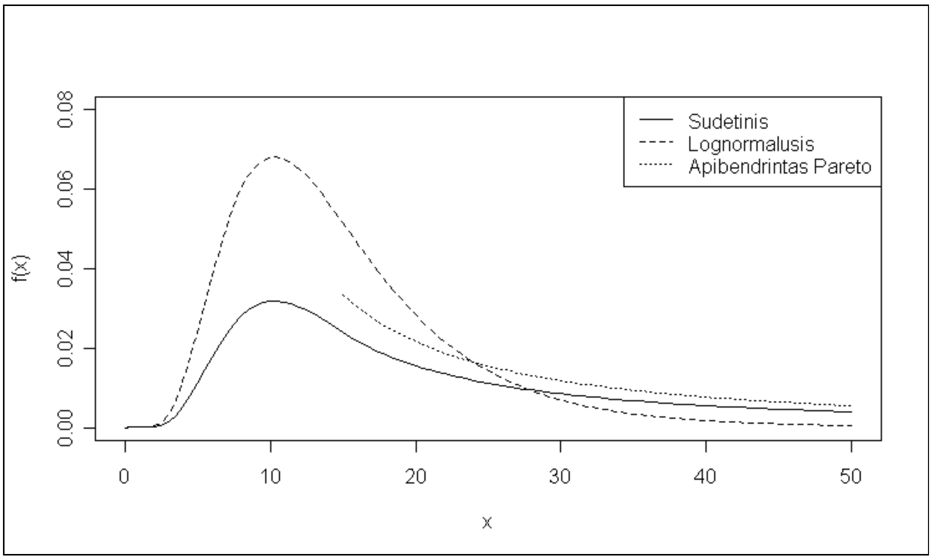
Gavome tolydų tankį priklausantį nuo keturių parametrų $\alpha > 0$, $\beta > 0$, $\theta > 0$, $\sigma > 0$. Iki sujungimo taško θ jis turi lognormaliojo tankio formą, o toliau tankio forma yra apibendrinto Pareto.

Statistiniame pakete „R“ sudėtinio lognormaliojo - apibendrinto Pareto skirstinio tankio funkciją galime aprašyti tokia funkcija[1]:

```
dlngpd=function(x)
   {ifelse (x>0&x<=t,
   (t/x*exp(-1/2*((log(x)-log(t)-s^2+((l+1)*t*s^2)/(l*b))/s)^2))/(b*exp(-
   1/2*(((l+1)*t*s)/(l*b)-s)^2)+pnorm(((l+1)*t*s)/(l*b)-s)*t*s*sqrt(2*pi)),
   exp(-1/2*(((l+1)*t*s)/(l*b)-s)^2)*(1+(x-t)/(l*b))^(-l-1)/(b*exp(-
   1/2*(((l+1)*t*s)/(l*b)-s)^2)+pnorm(((l+1)*t*s)/(l*b)-s)*t*s*sqrt(2*pi)))
   }
```

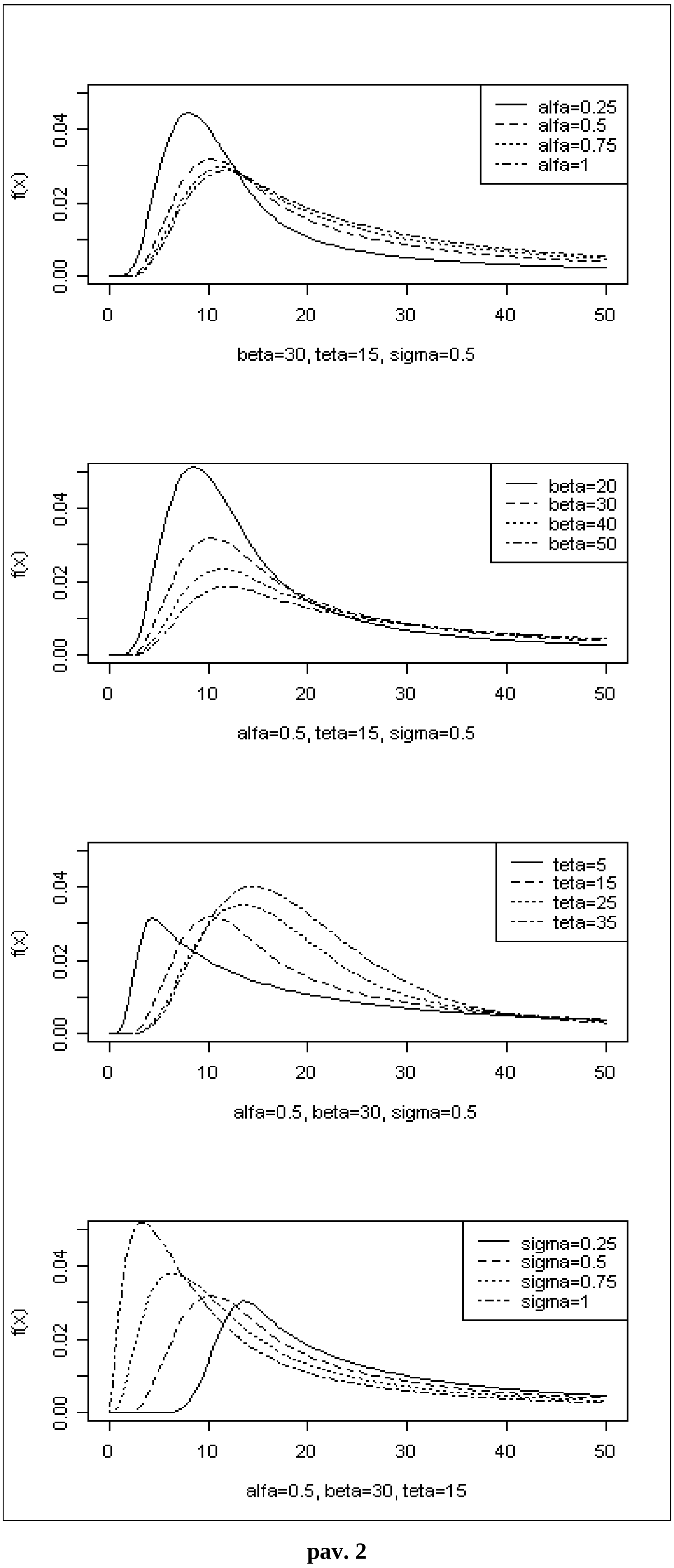
Pirmame paveikslėlyje pavaizduota kaip atrodo sudėtinis lognormalusis – apibendrintas Pareto skirstinys, lyginant jį su lognormaliuoju, bei apibendrintu Pareto. Jungiama taške θ = 15, parametro α reikšmė lygi 0,5, β = 30, σ = 0,5, o μ reikšmę galime susiskaičiuoti ($\mu=\sigma^2+\ln\theta-\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}$): μ = 2,58305. Taip pat galime paskaičiuoti kiekvienam iš šį modelį sudarančių skirstinių priskiriamą svorį: nupjautam lognormaliniam skirstiniui priskiriamas svoris d lygus 0,7209321, o apibendrintam Pareto - 1-d lygus 0,2790679. Iš grafiko galime matyti, jog lognormaliojo - apibendrinto Pareto tankis kinta nuo nulio ir iki taško θ įgyja lognormaliojo tankio formą, o nuo taško θ jau turi apibendrinto Pareto tankio formą, taigi, jo uodega yra gerokai sunkesnė nei lognormaliojo, tai yra, jis geriau aprašo dideles žalas.

[1] Šioje funkcijoje parametras α žymimas simboliu „l“, β - „b“, θ – „t“, σ - „s“.



pav. 1

Antrame paveikslėlyje pavaizduota, kaip kinta sudėtinio lognormaliojo - apibendrinto Pareto skirstinio tankis keičiant vieną iš parametrų:


alfa=0.25
alfa=0.5
alfa=0.75
alfa=1
f(x)
beta=30, teta=15, sigma=0.5
beta=20
beta=30
beta=40
beta=50
alfa=0.5, teta=15, sigma=0.5
teta=5
teta=15
teta=25
teta=35
alfa=0.5, beta=30, sigma=0.5
sigma=0.25
sigma=0.5
sigma=0.75
sigma=1
alfa=0.5, beta=30, teta=15

pav. 2

Paskaičiuotos svorių parametrų d ir $1\text{-}d$ reikšmės kiekvienam iš grafikuose naudojamų parametrų rinkinių:

$\beta=30;\ \theta=15;\ \sigma=0{,}5$			$\alpha=0{,}5;\ \theta=15;\ \sigma=0{,}5$			$\alpha=0{,}5;\ \beta=30;\ \sigma=0{,}5$			$\alpha=0{,}5;\ \beta=30;\ \theta=15$		
α	d	$1\text{-}d$	β	d	$1\text{-}d$	θ	d	$1\text{-}d$	σ	d	$1\text{-}d$
0,25	0,391	0,609	20	0,456	0,543	5	0,080	0,920	0,25	0,148	0,852
0,5	0,279	0,721	30	0,279	0,721	15	0,279	0,721	0,5	0,279	0,721
0,75	0,251	0,749	40	0,198	0,802	25	0,517	0,483	0,75	0,395	0,605
1	0,239	0,761	50	0,153	0,847	35	0,741	0,259	1	0,495	0,505

1 lentelė

Pirmoje lentelėje matome jog, didėjant parametrams α ir β, didėja apibendrinto Pareto skirstinio svoris mišinyje, o didinant parametrus θ ir σ, didiname lognormaliojo skirstinio proporciją.

1.2. Pasiskirstymo funkcija

Pasiskirstymo funkciją gausime integruodami sudėtinio lognormaliojo - apibendrinto Pareto skirstinio tankį nuo 0 iki x:

$$F(x) = \int_0^x f(s)ds.$$

Kadangi tankis iki taško θ ir už jo turi skirtingas išraiškas, tai pasiskirstymo funkcija turės tokią išraišką:

$$F(x) = \begin{cases} \displaystyle\int_0^x \frac{\frac{\theta}{s}\exp\left\{-\frac{1}{2}\left(\frac{\ln s - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)^2\right\} + \Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)\sqrt{2\pi}\theta\sigma}\,\mathrm{ds}, & \text{jei } 0 < x \le \theta; \\ \displaystyle\int_0^\theta \frac{\frac{\theta}{s}\exp\left\{-\frac{1}{2}\left(\frac{\ln s - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)^2\right\} + \Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)\sqrt{2\pi}\theta\sigma}\,\mathrm{ds} + \\ \quad \displaystyle\int_\theta^x \frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)^2\right\}\left(1 + \frac{s-\theta}{\alpha\beta}\right)^{-\alpha-1}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)^2\right\} + \Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)\sqrt{2\pi}\theta\sigma}ds, & \text{jei } \theta \le x < \infty. \end{cases}$$

Skaičiuojame pirmąjį integralą:

$$I_1 := \int_0^x \frac{\frac{\theta}{s}\exp\left\{-\frac{1}{2}\left(\frac{\ln s - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma} \mathrm{ds} =$$

$$\frac{\theta}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\int_0^x \frac{1}{s}\exp\left\{-\frac{1}{2}\left(\frac{\ln s - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\} ds;$$

Atliekame pakeitimą:

$$t = \frac{\ln s - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma};\ s = \exp\left\{t\sigma + \ln\theta + \sigma^2 - \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}\right\};\ ds = \sigma s dt\ .$$

Po pakeitimo turime:

$$I_1 := \frac{\theta}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\int_{-\infty}^{\frac{\ln x - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}} \frac{1}{s}\mathrm{e}^{-\frac{t^2}{2}}\sigma s dt =$$

$$\frac{\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\ln x - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}} \mathrm{e}^{-\frac{t^2}{2}} dt =$$

$$\frac{\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\Phi\left(\frac{\ln x - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right).$$

Antrosios dalies pirmasis integralas yra tas pats pirmasis integralas tik iki taško θ. Jis lygus:

$$\frac{\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma};$$

o antrąjį skaičiuojame taip:

$$I_2 := \int_{\theta}^{x} \frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}\left(1+\frac{s-\theta}{\alpha\beta}\right)^{-\alpha-1}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma} ds =$$

$$\frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\int_{\theta}^{x}\left(1+\frac{s-\theta}{\alpha\beta}\right)^{-\alpha-1} ds;$$

Atliekame pakeitimą: $t = 1+\frac{s-\theta}{\alpha\beta}; ds = \alpha\beta dt$:

$$I_2 := \frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\int_{1}^{1+\frac{x-\theta}{\alpha\beta}} t^{-\alpha-1}\alpha\beta dt =$$

$$\frac{\alpha\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\left(\frac{t^{-\alpha}}{-\alpha}\Bigg|_{1}^{1+\frac{x-\theta}{\alpha\beta}}\right) =$$

$$\frac{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\left(1-\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha}\right);$$

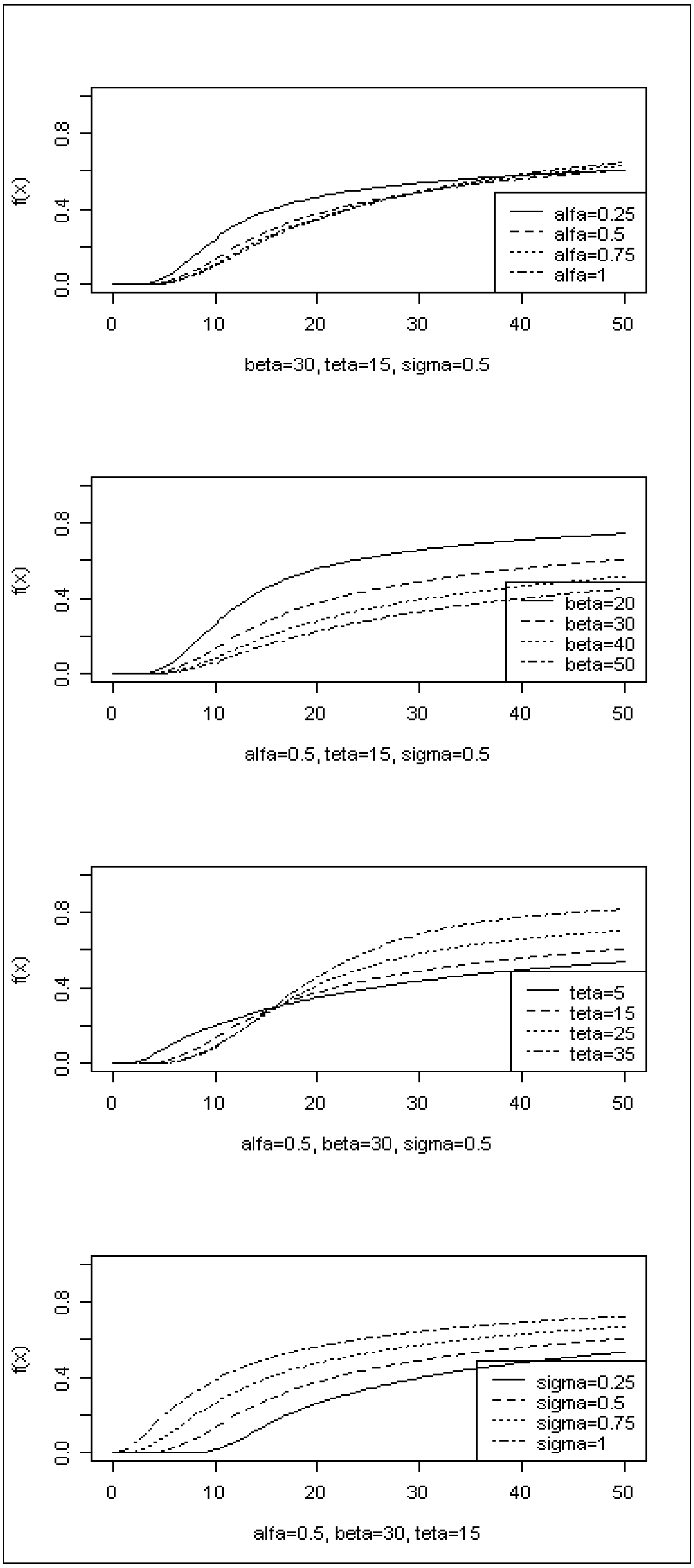
Turime tokią galutinę pasiskirstymo funkcijos išraišką:

$$F(x)=\begin{cases} \dfrac{\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\Phi\left(\dfrac{\ln x-\ln\theta-\sigma^2+\frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right), \text{ jei } 0<x\le\theta; \\ \dfrac{\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}+ \\ \dfrac{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\left(1-\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha}\right), \text{ jei } \theta\le x<\infty. \end{cases}$$

Statistiniame pakete „R“ sudėtinio lognormaliojo - apibendrinto Pareto skirstinio pasiskirstymo funkciją aprašome tokia funkcija:

```
plngpd=function(x)
{ifelse (x>0&x<=t,
   t*s*sqrt(2*pi)*pnorm((log(x)-log(t)-s^2+((l+1)*t*s^2)/(l*b))/s)/(b*exp(-
   1/2*(((l+1)*t*s)/(l*b)-s)^2)+pnorm(((l+1)*t*s)/(l*b)-s)*t*s*sqrt(2*pi)),
   (pnorm(((l+1)*t*s)/(l*b)-s)*t*s*sqrt(2*pi)+b*exp(-1/2*(((l+1)*t*s)/(l*b)-
   s)^2)*(1-(1+(x-t)/(l*b))^(-l)))/(b*exp(-1/2*(((l+1)*t*s)/(l*b)-
   s)^2)+pnorm(((l+1)*t*s)/(l*b)-s)*t*s*sqrt(2*pi)))
}
```

Trečiame paveikslėlyje pavaizduota, kaip keičiasi sudėtinio lognormaliojo - apibendrinto Pareto skirstinio pasiskirstymo funkcija keičiant vieną iš parametrų:


f(x)
alfa=0.25
alfa=0.5
alfa=0.75
alfa=1
beta=30, teta=15, sigma=0.5
beta=20
beta=30
beta=40
beta=50
alfa=0.5, teta=15, sigma=0.5
teta=5
teta=15
teta=25
teta=35
alfa=0.5, beta=30, sigma=0.5
sigma=0.25
sigma=0.5
sigma=0.75
sigma=1
alfa=0.5, beta=30, teta=15

pav. 3

1.3. t-tasis momentas

Rasime sudėtinio lognormaliojo – apibendrinto Pareto skirstinio t-tąjį momentą. Bendru atveju, t-tasis momentas randamas taip:

$$E(X^t)=\int_{-\infty}^{+\infty} x^t f(x)dx\,.$$

Šiuo atveju skaičiuosime dviejų integralų sumą:

$$E(X^t)=\int_0^{\theta} x^t \frac{\dfrac{\theta}{x}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}}{\beta\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}dx+$$

$$\int_{\theta}^{\infty} x^t \frac{\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}\left(1+\dfrac{x-\theta}{\alpha\beta}\right)^{-\alpha-1}}{\beta\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}dx;$$

Skaičiuojame pirmąjį integralą:

$$I_1:=\int_0^{\theta} x^t \frac{\dfrac{\theta}{x}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}}{\beta\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}dx=$$

$$\frac{\theta}{\beta\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\int_0^{\theta}\frac{x^t}{x}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}dx;$$

Pažymėję $c:=\dfrac{\theta}{\beta\exp\left\{-\dfrac{1}{2}\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\dfrac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}$, turime:

$$I_1 := c\int_0^{\theta}\frac{1}{x}\mathrm{e}^{t\ln x}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}dx = c\int_0^{\theta}\frac{1}{x}\exp\left\{t\ln x-\frac{1}{2}\left(\frac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}dx =$$

$$c\int_0^{\theta}\frac{1}{x}\exp\left\{\left(-\frac{1}{2}\left(\frac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2+\frac{1}{2}2\sigma t\left(\frac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)-\frac{1}{2}\sigma^2t^2\right)+\frac{1}{2}\sigma^2t^2-\sigma t\left(\frac{-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)\right\}dx =$$

$$c\int_0^{\theta}\frac{1}{x}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}-\sigma t\right)^2\right\}\exp\left\{\frac{1}{2}\sigma^2t^2-\sigma t\left(\frac{-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)\right\}dx =$$

$$c\exp\left\{\frac{1}{2}\sigma^2t^2-\sigma t\left(\frac{-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)\right\}\int_0^{\theta}\frac{1}{x}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}-\sigma t\right)^2\right\}dx;$$

Atlikime pakeitimą: $y := \dfrac{\ln x-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}-\sigma t;\quad dy=\dfrac{1}{\sigma x}dx;\quad dx=\sigma x dy;$

$$I_1 := c\exp\left\{\frac{1}{2}\sigma^2t^2-\sigma t\left(\frac{-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)\right\}\int_{-\infty}^{\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma-\sigma t}\frac{1}{x}\mathrm{e}^{-\frac{y^2}{2}}\sigma x dy =$$

$$c\exp\left\{\frac{1}{2}\sigma^2t^2-\sigma t\left(\frac{-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)\right\}\sigma\sqrt{2\pi}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma-\sigma t}\mathrm{e}^{-\frac{y^2}{2}}dx =$$

$$c\exp\left\{\frac{1}{2}\sigma^2t^2-\sigma t\left(\frac{-\ln\theta-\sigma^2+\dfrac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)\right\}\sigma\sqrt{2\pi}\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma-\sigma t\right);$$

Grąžinę pakeistąją konstantą c gauname tokią pirmojo integralo išraišką:

$$\frac{\exp\left\{\frac{1}{2}\sigma^2t^2-t\left(-\ln\theta-\sigma^2+\frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}\right)\right\}\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma-\sigma t\right)$$

Dabar skaičiuosime antrąjį integralą:

$$I_2:=\int_\theta^\infty x^t\frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha-1}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}dx=$$

$$\frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\int_\theta^\infty x^t\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha-1}dx$$

Pažymėkime: $c:=\dfrac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}$;

Taigi mums reikia skaičiuoti tokį integralą: $c\int_\theta^\infty x^t\left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha-1}dx$. Kadangi šio integralo skaičiavimas yra gana komplikuotas, jo išraiškai gauti panaudosime kompiuterinę programą skirtą matematiniams skaičiavimams “Maple” (versija 10.00). Į “Maple” konsolę įvedę tokį kodą:

```
c*int(x^t*(1+(x-theta)/(alpha*beta))^(-alpha-1),x=theta..infinity, assume(x>0,
alpha>0, beta>0, theta>0));
```

gausime integralo išraišką:

$$c\frac{\left(\frac{1}{\alpha\beta-\theta}\right)^{-\alpha}\left(\frac{\alpha\beta-\theta}{\alpha\beta}\right)^{-\alpha-1}(\alpha\beta-\theta)\theta^{t-\alpha}\,{}_2F_1\left(-t+\alpha,\alpha+1;1-t+\alpha;-\frac{\alpha\beta-\theta}{\theta}\right)}{-t+\alpha}=$$

$$c\frac{(\alpha\beta)^{\alpha+1}\theta^{t-\alpha}\,{}_2F_1\left(-t+\alpha,\alpha+1;1-t+\alpha;-\frac{\alpha\beta-\theta}{\theta}\right)}{-t+\alpha};$$

Čia ${}_2F_1(a,b;c;z)$ yra hipergeometrinė eilutė, jos išraiška yra:

$$ {}_2F_1\left(a,b;c;z\right)=\sum_{n=0}^{\infty}\frac{(a)_n\left(b\right)_n}{\left(c\right)_n}\frac{z^n}{n!}; \quad \left(a\right)_n=a\left(a+1\right)\left(a+2\right)\ldots\left(a+n-1\right); $$

Ši eilutė konverguoja, kai $c\geq 0$ visiems $\left|z\right|<1$.

$$ {}_2F_1\left(-t+\alpha,\alpha+1;1-t+\alpha;-\frac{\alpha\beta-\theta}{\theta}\right)=\sum_{n=0}^{\infty}\frac{(-t+\alpha)_n\left(\alpha+1\right)_n}{\left(1-t+\alpha\right)_n}\frac{\left(-\frac{\alpha\beta-\theta}{\theta}\right)^n}{n!}= $$

$$ \sum_{n=0}^{\infty}\frac{(-t+\alpha)(-t+\alpha+1)(-t+\alpha+2)...(-t+\alpha+n-1)\left(\alpha+1\right)_n}{\left(1-t+\alpha\right)\left(1-t+\alpha+1\right)...\left(1-t+\alpha+n-2\right)\left(1-t+\alpha+n-1\right)}\frac{\left(-\frac{\alpha\beta-\theta}{\theta}\right)^n}{n!}= $$

$$ \sum_{n=0}^{\infty}\frac{(-t+\alpha)\left(\alpha+1\right)_n}{\left(-t+\alpha+n\right)}\frac{\left(-\frac{\alpha\beta-\theta}{\theta}\right)^n}{n!}=(-t+\alpha)\sum_{n=0}^{\infty}\frac{\left(\alpha+1\right)_n}{\left(-t+\alpha+n\right)}\frac{\left(-\frac{\alpha\beta-\theta}{\theta}\right)^n}{n!}; $$

Po kai kurių prastinimų bei gražinę parametro c išraišką, galiausiai turime tokią t-tojo momento išraišką:

$$ E(X^t)=\frac{\exp\left\{\frac{1}{2}\sigma^2t^2-t\left(-\ln\theta-\sigma^2+\frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}\right)\right\}\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma-\sigma t\right)+ $$

$$ \frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}\left(\alpha\beta\right)^{\alpha+1}\theta^{t-\alpha}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\sum_{n=0}^{\infty}\frac{\left(\alpha+1\right)_n}{\left(-t+\alpha+n\right)}\frac{\left(-\frac{\alpha\beta-\theta}{\theta}\right)^n}{n!}. $$

Kad eilutė $\sum_{n=0}^{\infty}\frac{\left(\alpha+1\right)_n}{\left(-t+\alpha+n\right)}\frac{\left(-\frac{\alpha\beta-\theta}{\theta}\right)^n}{n!}$ konverguotų, turi galioti nelygybė $1-t+\alpha\geq 0$ visiems $\left|-\frac{\alpha\beta-\theta}{\theta}\right|<1$, arba, $1+\alpha\geq t$ visiems $\alpha\beta<2\theta$.

Kai t reikšmės yra natūrinės, integralą $I := \int_{\theta}^{\infty} x^t \left(1+\frac{x-\theta}{\alpha\beta}\right)^{-\alpha-1} dx$ skaičiuoti nebėra taip sudėtinga, tereikia pritaikyti tokį pakeitimą: $y := 1+\frac{x-\theta}{\alpha\beta};\ x = (y-1)\alpha\beta+\theta;\ dx = \alpha\beta dy.$ Tuomet galime taikyti Niutono binomo formulę:

$$I := \alpha\beta\int_1^{\infty}(\alpha\beta)^t\left(y+\frac{\theta}{\alpha\beta}-1\right)^t y^{-\alpha-1}dy = (\alpha\beta)^{t+1}\int_1^{\infty}\sum_{k=0}^{t} C_t^k y^k\left(\frac{\theta}{\alpha\beta}-1\right)^{t-k} y^{-\alpha-1}dy =$$

$$(\alpha\beta)^{t+1}\sum_{k=0}^{t}\int_1^{\infty} C_t^k y^{k-\alpha-1}\left(\frac{\theta}{\alpha\beta}-1\right)^{t-k} dy = (\alpha\beta)^{t+1}\sum_{k=0}^{t} C_t^k\left(\frac{\theta}{\alpha\beta}-1\right)^{t-k}\int_1^{\infty} y^{k-\alpha-1}dy =$$

$$(\alpha\beta)^{t+1}\sum_{k=0}^{t} C_t^k\left(\frac{\theta}{\alpha\beta}-1\right)^{t-k}\left(\left.\frac{y^{k-\alpha}}{k-\alpha}\right|_1^{\infty}\right) = (\alpha\beta)^{t+1}\sum_{k=0}^{t} C_t^k\left(\frac{\theta}{\alpha\beta}-1\right)^{t-k}\left(\frac{\lim\limits_{y\to\infty} y^{k-\alpha}-1}{k-\alpha}\right);$$

Ši eilutė konverguos tik kai $k < \alpha, \forall k = 0..t$, tai yra, kai $t < \alpha$. Tuomet turėsime:

$$I := (\alpha\beta)^{t+1}\sum_{k=0}^{t} C_t^k\left(\frac{\theta}{\alpha\beta}-1\right)^{t-k}\left(\frac{1}{\alpha-k}\right).$$

Šiuo atveju, kai t – natūriniai skaičiai, momentų išraiška atrodo taip:

$$E(X^t) = \frac{\exp\left\{\frac{1}{2}\sigma^2 t^2 - t\left(-\ln\theta-\sigma^2+\frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}\right)\right\}\sqrt{2\pi}\theta\sigma}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma-\sigma t\right)+$$

$$\frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}(\alpha\beta)^{t+1}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}+\Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma}\sum_{k=0}^{t} C_t^k\left(\frac{\theta}{\alpha\beta}-1\right)^{t-k}\left(\frac{1}{\alpha-k}\right).$$

2. PARAMETRŲ ĮVERČIAI

2.1. Tikėtinumo funkcija

Sakykime, $X_1, X_2,..., X_n$ yra imtis iš sudėtinio lognormaliojo - apibendrinto Pareto skirstinio su keturiais parametrais. Nemažindami bendrumo, galime laikyti, kad imtis sutvarkyta, t.y. $x_1 \le x_2 \le x_3 \le ... \le x_n$. Tarkime, nežinomas parametras θ yra tarp m-tojo ir (m+1)-ojo stebėjimo, t.y. $x_m \le \theta \le x_{m+1}$. Tikėtinumo funkciją šiam skirstiniui bendrai galima užrašyti taip:

$$L(\alpha,\beta,\theta,\sigma) = f(x_1,\alpha,\beta,\theta,\sigma)...f(x_n,\alpha,\beta,\theta,\sigma);$$

Kadangi iki taško θ tankio funkcija turi vieną išraišką, o už jo - kitą, tai iki m-tojo stebėjimo mes turime imti ir dauginti pirmąją tankio funkcijos išraišką, o nuo (m+1)-ojo – antrąją. Taigi skaičiuojame:

$$L(\alpha,\beta,\theta,\sigma) = \prod_{i=1}^{m} \frac{\frac{\theta}{x_i}\exp\left\{-\frac{1}{2}\left(\frac{\ln x_i - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\} + \Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma} \times$$

$$\prod_{i=m+1}^{n} \frac{\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}\left(1+\frac{x_i-\theta}{\alpha\beta}\right)^{-\alpha-1}}{\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\} + \Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma} =$$

$$\left(\beta\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\} + \Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)\sqrt{2\pi}\theta\sigma\right)^{-n} \prod_{i=1}^{m}\frac{\theta}{x_i} \times$$

$$\prod_{i=1}^{m}\exp\left\{-\frac{1}{2}\left(\frac{\ln x_i - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\}\left(\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta}-\sigma\right)^2\right\}\right)^{n-m}\prod_{i=m+1}^{n}\left(1+\frac{x_i-\theta}{\alpha\beta}\right)^{-\alpha-1};$$

Skaičiuokime tikėtinumo funkcijos natūrinį logaritmą:

$$\Lambda = \ln L\left(\alpha, \beta, \theta, \sigma\right) = -n\ln\left(\beta \exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)^2\right\} + \Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)\sqrt{2\pi}\theta\sigma\right) + \ln\prod_{i=1}^{m}\frac{\theta}{x_i} +$$

$$\ln\prod_{i=1}^{m}\exp\left\{-\frac{1}{2}\left(\frac{\ln x_i - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\} + (n-m)\ln\left(\exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)^2\right\}\right) +$$

$$\ln\prod_{i=m+1}^{n}\left(1 + \frac{x_i - \theta}{\alpha\beta}\right)^{-\alpha-1};$$

Po tam tikrų pertvarkymų turime galutinę logaritminės tikėtinumo funkcijos išraišką:

$$\Lambda = -n\ln\left(\beta \exp\left\{-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)^2\right\} + \Phi\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)\sqrt{2\pi}\theta\sigma\right) + \sum_{i=1}^{m}\ln\frac{\theta}{x_i} +$$

$$\sum_{i=1}^{m}\left\{-\frac{1}{2}\left(\frac{\ln x_i - \ln\theta - \sigma^2 + \frac{(\alpha+1)\theta\sigma^2}{\alpha\beta}}{\sigma}\right)^2\right\} + (n-m)\left(-\frac{1}{2}\left(\frac{(\alpha+1)\theta\sigma}{\alpha\beta} - \sigma\right)^2\right) + \sum_{i=m+1}^{n}\ln\left(1 + \frac{x_i - \theta}{\alpha\beta}\right).$$

Kadangi ši logaritminė tikėtinumo funkcija yra ganėtinai sudėtinga, lygčių sistemos, sudarytos iš dalinių išvestinių prilygintų nuliui, sprendimas yra komplikuotas, analiziniu būdu rasti didžiausio tikėtinumo įverčius yra beveik neįmanoma. Todėl ieškodami įverčių naudosimės viena iš statistiniame pakete „R“ aprašytų funkcijų, didžiausio tikėtinumo įverčiams rasti.

2.2. Įverčių radimas

Kaip minėta 3.1. skyrelyje, įverčių radimo procedūra yra gana komplikuota ir, dėl sudėtingos logaritminės didžiausio tikėtinumo funkcijos išraiškos, įverčiai skaičiuojami artutiniais metodais. Tam naudojame vieną iš statistinės programos „R“ funkcijų. Funkcija „maxLik“[2] (esanti pakete „maxLik“) ir yra skirta tikimybinių tankių įverčių radimo problemai spręsti. Įverčių skaičiavimui naudojamas Nelder - Mead minimizavimo metodas.

Nelder - Mead metodas – simplekso metodas skirtas rasti kelių kintamųjų funkcijos lokaliems minimumams. Jis pavadintas kūrėjų J. A. Nelder ir R. Mead (1965) garbei. Metodas yra gana patogus, kadangi nereikalauja daug skaičiavimų, taigi yra ir efektyvus. Esant didelėms imtims tai yra gana svarbu, nes, atsižvelgiant į tai, jog turint n dydžio tūrį, funkciją tenka maksimizuoti n-1 kartą (t.y., laikant, kad sujungimo taškas yra tarp m-tojo ir (m+1)-ojo surikiuotos imties nario, maksimizuojama kiekvienam m nuo 1 iki n-1), skaičiavimai kompiuteriu trunka gana ilgai.

Kita su maksimizavimu susijusi problema yra ta, kad logaritminėje didžiausio tikėtinumo funkcijoje esantys logaritmai, esant tam tikroms parametrų reikšmėms gali būti neapibrėžti, t.y. po logaritmo ženklu esančiai funkcijai įgijus neteigiamą reikšmę, logaritmo neįmanoma suskaičiuoti. Taigi ir maksimizavimo metodas (kai kuriems m) gali duoti nepageidautinus rezultatus (pvz. nepagrįstai didelę logaritminės didžiausio tikėtinumo funkcijos reikšmę). Vis dėlto empiriškai matosi[3], kad parametrų reikšmėms esant artimoms tikrosioms metodas duoda gerus rezultatus. Svarbu pažymėti, kad maksimizavimo rezultatams reikšmės turi pradinių parametrų pasirinkimas. Jas galima pabandyti įvertinti empiriškai, pabandžius nusibraižyti duomenų histogramą ir pamėginus tame pačiame grafike nusibraižyti tankį, parametrų reikšmes pasirenkant taip, kad tiek histograma, tiek tankis turėtų panašią formą.

Iš visų įverčių, gautų didžiausio tikėtinumo metodu, kurių turime n-1 išsirinkti geriausią galime ieškodami mažiausio atstumo tarp skirstinio ir duomenų imties. Tokia atstumo funkcija galime laikyti vieną iš testų, skirtų tikrinti hipotezėms apie duomenų atitikimą skirstiniui (duomenų analizėje naudojami Kolmogorovo – Smirnovo, Anderseno – Darlingo testai, Chi kvadrato suderinamumo kriterijus). Toliau pateiktoje „R“ programai skirtoje funkcijoje naudosime Kolmogorovo – Smirnovo, Anderseno – Darlingo testus (Chi kvadrato suderinamumo kriterijus nėra patogus, kadangi jo reikšmės priklauso nuo to, į kokius intervalus padalinta imtis).

[2] Ši funkcija suderinama tik su 2.6.2 ir naujesnėmis „R“ versijomis.
[3] Duomenų analizės rezultatai aptariami trečiame skyriuje.

Teoriškai, gautas sujungimo taško θ įvertis turėtų būti tarp m-tojo bei (m+1)-ojo variacinės eilutės nario, tačiau, atsižvelgiant į tai, kad įverčiai randami artutiniais metodais, galime neišvengti įverčių paslinktumo.

Toliau pateikiamas įverčių radimo funkcijos kodas. Pradiniai funkcijos „lngpdmle“ įvesties parametrai yra (iš eilės) duomenų vektorius, pradinės parametrų α, β, θ, σ reikšmės:

```
lngpdmle=function(xx,a,b,c,d) {
   x=sort(xx) # imties variacine eilute
   n=length(x)
   ll=a
   bb=b
   tt=c
   ss=d
   ltemp=0
   btemp=0
   ttemp=0
   stemp=0
   dtf=0
   l=0
   b=0
   t=0
   s=0
   plngpd=function(x) #apsirasome pasiskirstymo funkcija
   {ifelse (x>0&x<=t,
      t*s*sqrt(2*pi)*pnorm((log(x)-log(t)-s^2+((l+1)*t*s^2)/(l*b))/s)/(b*exp(-
      1/2*(((l+1)*t*s)/(l*b)-s)^2)+pnorm(((l+1)*t*s)/(l*b)-s)*t*s*sqrt(2*pi)),
      (pnorm(((l+1)*t*s)/(l*b)-s)*t*s*sqrt(2*pi)+b*exp(-
      1/2*(((l+1)*t*s)/(l*b)-s)^2)*(1-(1+(x-t)/(l*b))^(-l)))/(b*exp(-
      1/2*(((l+1)*t*s)/(l*b)-s)^2)+pnorm(((l+1)*t*s)/(l*b)-s)*t*s*sqrt(2*pi)))
   }
   kstestas=function(z,f){ #apsirasome Kolmogorovo-Smirnovo testo algoritma
      n=length(z)
      y=sort(z)
      funcc=f
      dpl=0
      for(i in 1:n) dpl[i]=i/n-funcc(y[i])
      dp=max(dpl)
      dmn=0
      for(i in 1:n) dmn[i]=funcc(y[i])-(i-1)/n
```

```
   dm=max(dmn)
   ks=max(dp,dm)
   return(ks)
}
adtestas=function(z,f){ #apsirasome Anderseno-Darlingo testo algoritma
   n=length(z)
   y=sort(z)
   funcc=f
   ss=0
   for(i in 1:n){ss=ss+(((2*i-1)/n)*(log(funcc(y[i]))+log(1-funcc(y[n+1-
   i]))))}
   return(-ss-n)
}
for (m in 1:(n-1)){ #Skaiciuojame didziausio tiketinumo ivercius
kiekviename taske
tik=function(param){
l<-param[1]
b<-param[2]
t<-param[3]
s<-param[4]
sum1=0
sum2=0
sum3=0
for(i in 1:m) sum1=sum1+log(t/x[i])
for(i in 1:m) sum2=sum2+(-1/2*((log(x[i])-log(t)-
s^2+((l+1)*t*s^2)/(l*b))/s)^2)
for(i in (m+1):n) sum3=sum3+log(1+(x[i]-t)/(l*b))
aa<-(-n)*log(b*exp(-1/2*(((l+1)*t*s)/(l*b)-s)^2)+pnorm(((l+1)*t*s)/(l*b)-
s)*t*s*sqrt(2*pi))+sum1+sum2+(n-m)*(-1/2*(((l+1)*t*s)/(l*b)-s)^2)-
(l+1)*sum3
}
tempfunc= maxLik(tik, start=c(ll,bb,tt,ss),method = "NM")
ltemp[m]= tempfunc $estimate[1]
btemp[m]= tempfunc $estimate[2]
ttemp[m]= tempfunc $estimate[3]
stemp[m]= tempfunc $estimate[4]
dtf[m]=tik(c(ltemp[m],btemp[m],ttemp[m],stemp[m]))
}
lngpdks=0
```

```
    for (i in 1:(n-1)){ #Randame ivercius remdamiesi Kolmogorovo-Smirnovo
    testo statistika
    l=ltemp[i]
    b=btemp[i]
    t=ttemp[i]
    s=stemp[i]
    lngpdks[i]=kstestas(x,plngpd)
  }
  whichminks=which(lngpdks==min(lngpdks))
  minks=lngpdks[whichminks]
  lmlks=ltemp[which(lngpdks==min(lngpdks))]
  bmlks=btemp[which(lngpdks==min(lngpdks))]
  tmlks=ttemp[which(lngpdks==min(lngpdks))]
  smlks=stemp[which(lngpdks==min(lngpdks))]
  lngpdad=0
    for (i in 1:(n-1)){ #Randame ivercius remdamiesi Anderseno-Darlingo
    testo statistika
    l=ltemp[i]
    b=btemp[i]
    t=ttemp[i]
    s=stemp[i]
    lngpdad[i]=adtestas(x,plngpd)
  }
  whichminad=which(lngpdad==min(lngpdad))
  minad=lngpdad[whichminad]
  lmlad=ltemp[which(lngpdad==min(lngpdad))]
  bmlad=btemp[which(lngpdad==min(lngpdad))]
  tmlad=ttemp[which(lngpdad==min(lngpdad))]
  smlad=stemp[which(lngpdad==min(lngpdad))]

  return(ltemp,btemp,ttemp,stemp,dtf,lmlks,bmlks,tmlks,smlks,whichminks,minks
  ,lmlad,bmlad,tmlad,smlad,whichminad,minad)
}
```

Kadangi rezultatų analizei mus gali dominti ne vien gauti parametrai, bet ir kiti skaičiavimo rezultatai, funkcija grąžina kai kuriuos tarpinius rezultatus:

ltemp – parametro α įverčių vektorius;

btemp – parametro β įverčių vektorius;

ttemp – parametro θ įverčių vektorius;

stemp – parametro σ įverčių vektorius;

dtf – logaritminės didžiausio tikėtinumo funkcijos su atitinkamais parametrų įverčiais reikšmių vektorius;

lmlks - parametro α įvertis, gautas remiantis K-S testo statistika;

bmlks - parametro β įvertis, gautas remiantis K-S testo statistika;

tmlks - parametro θ įvertis, gautas remiantis K-S testo statistika;

smlks - parametro σ įvertis, gautas remiantis K-S testo statistika;

whichminks – mažiausios K-S testo reikšmės vektoriaus koordinatė;

minks – mažiausia K-S testo reikšmė;

lmlad - parametro α įvertis, gautas remiantis A-D testo statistika;

bmlad - parametro β įvertis, gautas remiantis A-D testo statistika;

tmlad - parametro θ įvertis, gautas remiantis A-D testo statistika;

smlad - parametro σ įvertis, gautas remiantis A-D testo statistika;

whichminad – mažiausios A-D testo reikšmės vektoriaus koordinatė;

minad – mažiausia A-D testo reikšmė;

3. DUOMENŲ ANALIZĖ

Šiame skyriuje nagrinėsime tris duomenų imtis: Kahadawala Cooray, Malwane M. A. Amanda straipsnyje „Modeling actuarial data with a composite lognormal – Pareto model“ generuotą ir nagrinėtą 50 atsitiktinių dydžių imtį, Danijos gaisrų draudimo žalas (2492 tūrio imtis), vienos Lietuvos draudimo kompanijos ne gyvybės draudimo žalas (809 tūrio dydžio imtis). Be šiame darbe nagrinėjamo skirstinio, palyginimui naudosime dar keletą skirstinių:

- jau minėtame straipsnyje „Modeling actuarial data with a composite lognormal – Pareto model“ nagrinėtą sudėtinį lognormalųjį – Pareto skirstinį:

$$f(x) = \begin{cases} \dfrac{\alpha\theta^{\alpha}}{(1+\Phi(k))x^{\alpha+1}} \exp\left\{ -\dfrac{\alpha^2}{2k^2} \ln^2(x/\theta) \right\}, & \text{jei } 0 < x \le \theta \\ \dfrac{\alpha\theta^{\alpha}}{(1+\Phi(k))x^{\alpha+1}}, & \text{jei } \theta \le x < \infty. \end{cases}$$

  su parametrais $\alpha > 0$ ir $\theta > 0$;

- lognormalųjį, kurio tankio funkcija:

$$f(x) = \frac{(2\pi)^{-1/2}}{x\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma} \right)^2 \right\}, x > 0$$

  su parametrais $\sigma > 0$ ir $\mu > 0$;

- Pareto, kurio tankio funkcija

$$f(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, x > \theta$$

  su parametrais $\alpha > 0$ ir $\theta > 0$;

- gama, kurio tankio funkcija

$$f(x) = \frac{\theta^{-\alpha} x^{\alpha-1}}{\Gamma(\alpha)} \exp\left\{ -\frac{x}{\theta} \right\}$$

  su parametrais $\alpha > 0$ ir $\theta > 0$. Čia $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$;

- Weibull, kurio tankio funkcija

$$f(x) = \left(\frac{\alpha}{x} \right)\left(\frac{x}{\theta} \right)^{\alpha} \exp\left\{ -\left(\frac{x}{\theta} \right)^{\alpha} \right\}$$

  su parametrais $\alpha > 0$ ir $\theta > 0$.

Remiantis Kahadawala Cooray ir Malwane M. A. Amanda straipsniu, V. Čekanavičiaus, G. Murausko knyga „Statistika ir jos taikymai“ bei J. Kruopio knyga „Matematinė statistika“, pasirinkti šie testai:

- Kolmogorovo - Smirnovo testo statistika (K- S). Šis testas yra naudojamas patikrinti, ar imtis yra pasiskirsčiusi pagal norimą skirstinį. Testo statistika D_n apskaičiuojama taip:

$$D_n = \max_{1 \le i \le n}\left(\frac{i}{n} - F(x_i), F(x_i) - \frac{i-1}{n}\right),$$

kur n - imties dydis, o $F(x_i)$ - pasiskirstymo funkcijos reikšmė taške x_i. Kuo statistikos D_n reikšmė mažesnė, tuo pasirinktas skirstinys yra geresnis. Statistiniame pakete „R“ šio testo statistikai suskaičiuoti galime pasinaudoti tokia funkcija:

```
kstestas=function(z,f){
   n=length(z)
   y=sort(z)
   funcc=f
   dpl=0
   for(i in 1:n) dpl[i]=i/n-funcc(y[i])
   dp=max(dpl)
   dmn=0
   for(i in 1:n) dmn[i]=funcc(y[i])-(i-1)/n
   dm=max(dmn)
   ks=max(dp,dm)
   return(ks)
}
```

čia `z` - empirinė imtis, `f` - teorinė pasiskirstymo funkcija;

- Anderseno - Darlingo testo statistika (A - D). Šis testas yra K- S testo atmaina ir taip pat naudojamas patikrinti, ar imtis yra pasiskirsčiusi pagal norimą skirstinį, tačiau jis yra tikslesnis. Testo statistika A^2 apskaičiuojama taip:

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n}\left(\ln(F(x_i)) + \ln(1 - F(x_{n+1-i}))\right),$$

kur n - imties dydis, o $F(x_i)$ - pasiskirstymo funkcijos reikšmė taške x_i. Kuo statistikos A^2 reikšmė mažesnė, tuo pasirinktas skirstinys geresnis. Statistiniame pakete „R“ šio testo statistikai suskaičiuoti naudosimės šia funkcija:

```
adtestas=function(z,f){
   n=length(z)
   y=sort(z)
```

```
   funcc=f
   ss=0
   for(i in 1:n){ss=ss+(((2*i-1)/n)*(log(funcc(y[i]))+log(1-funcc(y[n+1-
i]))))}
   return(-ss-n)
}
```

čia `z` - empirinė imtis, `f` - teorinė pasiskirstymo funkcija;

- Chi kvadrato suderinamumo kriterijaus p - reikšmė. Šis kriterijus yra alternatyva K - S ir A - D testams. Jis yra taikomas sugrupuotiems duomenims ir jo reikšmė priklauso nuo pasirinktų grupavimo intervalų. Tarkim, sutvarkyta imtis yra suskirstyta į k intervalų. Tuomet tikimybė, kad imties reikšmė pateks į i-tąjį intervalą apskaičiuojama taip:

$$p_i = F(x_i) - F(x_{i-1}), i = 1,...,k.$$

Čia $F(x_i)$ - pasiskirstymo funkcijos reikšmė taške x_i.

Testo statistika X^2 apskaičiuojama taip:

$$X^2 = \sum_{i=1}^{k} \frac{(V_i - np_i)^2}{np_i},$$

kur n - imties dydis, V_i - imties reikšmių, patekusių į i-tąjį intervalą, skaičius.

p - reikšmė bus lygi tikimybei $P\{\chi^2_{k-s-1} > X^2\}$, kur s - nežinomų parametrų skaičius, o χ^2_{k-s-1} - chi kvadrato pasiskirstymo funkcija su $k-s-1$ laisvės laipsniais. Kuo ši reikšmė didesnė, tuo pasirinktas skirstinys geresnis. Statistiniame pakete „R“ šio testo statistikai suskaičiuoti naudosimės šia funkcija:

```
chitest=function(z,f,aa,bb,cc,dd,ee,ff){
xsrt=sort(z)
   x1=0
   xq=0
   xq[1]=1
   for(i in 1:length(x))
   {
   if(xsrt[i]<aa) x1[i]=xsrt[i]
   }
   xq[2]=length(x1)
   x1=0
   for(i in 1:length(x))
   {
   if(xsrt[i]<bb) x1[i]=xsrt[i]
```

```
xq[3]=length(x1)
}
x1=0
for(i in 1:length(x))
{
if(xsrt[i]<cc) x1[i]=xsrt[i]
xq[4]=length(x1)
}
x1=0
for(i in 1:length(x))
{
if(xsrt[i]<dd) x1[i]=xsrt[i]
xq[5]=length(x1)
}
n=length(xsrt)
x1=0
for(i in 1:length(x))
{
if(xsrt[i]<ee) x1[i]=xsrt[i]
xq[6]=length(x1)
}
x1=0
for(i in 1:length(x))
{
if(xsrt[i]<ff) x1[i]=xsrt[i]
xq[7]=length(x1)
}
xq[8]=length(xsrt)
xf=0
xf[1]=xq[1]
xf[2]=xq[2]
for(i in 1:6)
{
xf[i+2]=xq[i+2]-xq[i+1]
}
funcc=f
ss=0
e=0
for(i in 1:7) {
e[i]=n*(funcc(xsrt[xq[i+1]])-funcc(xsrt[xq[i]]))
```

```
  ss=ss+(xf[i+1]-e[i])^2/e[i]
  }
  return(ss)
}
```
čia `z` - empirinė imtis, `f` - teorinė pasiskirstymo funkcija. Pastarojoje funkcijoje reikia nurodyti intervalų, į kuriuos skaidome turimą imtį, galus (funkcijoje šie parametrai pažymėti `aa`, `bb`, `cc`, `dd`, `ee`, `ff`).

3.1. Sudėtiniu lognormaliuoju - Pareto skirstiniu sugeneruotų duomenų analizė

Šiame skyrelyje analizuosime Kahadawala Cooray, Malwane M. A. Amanda straipsnyje „Modeling actuarial data with a composite lognormal – Pareto model“ nagrinėtus duomenis, sugeneruotus lognormaliuoju - Pareto skirstiniu. Imtį sudaro 50 atsitiktinių dydžių. Taip atrodo šios imties variacinė eilutė:

306, 383, 395, 450, 503, 595, 631, 663, 684, 717, 814, 884, 898, 903, 907, 937, 958, 968, 979, 1056, 1137, 1284, 1361, 1368, 1382, 1394, 1525, 1606, 1738, 1966, 2110, 2295, 2407, 2430, 2630, 3199, 3287, 3338, 3398, 3985, 4711, 6193, 6538, 7171, 8766, 9366, 20393, 58263, 59166, 400976.

3.1.1 Skaičiavimai

- Pirmiausia nagrinėkime sudėtinį lognormalųjį - apibendrintą Pareto modelį. Su antrame skyriuje aprašyta funkcija įverčiams gauti, galime rasti nežinomų parametrų įverčius. Programoje „R“ turimą imtį apsirašykime vektoriumi (šiuo atveju mes jį pažymėsime simboliu „x“). Paleidę įverčių skaičiavimo funkciją „lngpdmle“:

```
lngpdmle(x,1,1000,1000,1);
```

gauname tokius rezultatus: Kolmogorovo – Smirnovo testu išrinkti įverčiai: $\hat{\alpha}_{KS}$=0,7327476, $\hat{\beta}_{KS}$=1716,708, $\hat{\theta}_{KS}$=1400,044, $\hat{\sigma}_{KS}$=0,5766252; Anderseno – Darlingo testu išrinkti įverčiai: $\hat{\alpha}_{AD}$=0,7233132, $\hat{\beta}_{AD}$=1773,036, $\hat{\theta}_{AD}$=1473,014, $\hat{\sigma}_{AD}$=0,5885831. Užsirašę pirmame skyriuje pateiktą pasiskirstymo funkciją, panaudodami šio skyriaus pradžioje užrašytas testų funkcijas, suskaičiuokime testų statistikų reikšmes turimiems parametrų įverčiams[4]:

```
l=0.7327476
b=1716.708
t=1400.044
s=0.5766252
kstestas(x,plngpd)
adtestas(x,plngpd)
chitest(x,plbts,650,950,1200,1700,3000,7000)
```

[4] Šiame skyriuje nagrinėjamose „R“ funkcijose parametras α žymimas simboliu „a“ arba „l“, β - „b“, θ – „t“, σ - „s“, μ –„m“.

Pirmiesiems įverčiams gauname tokias statistikų reikšmes: Kolmogorovo – Smirnovo testo: 0,05603657; Anderseno – Darlingo testo: 0,1298399. Gavę chi kvadrato suderinamumo kriterijaus reikšmę, iš lentelių ar pasinaudoję specialiomis skaičiuoklėmis randame p reikšmę: 0,8853. Dabar suskaičiuokime šių testų statistikas antriesiems įverčiams:

```
l=0.7327476
b=1716.708
t=1400.044
s=0.5766252
kstestas(x,plngpd)
adtestas(x,plngpd)
chitest(x,plbts,650,950,1200,1700,3000,7000)
```

Gauname tokias statistikų reikšmes: Kolmogorovo – Smirnovo testo: 0,05833153; Anderseno – Darlingo testo: 0,1292431; chi kvadrato suderinamumo kriterijaus p reikšmė: 0,8633.

- Dabar nagrinėkime lognormalųjį – Pareto skirstinį. Jo didžiausio tikėtinumo parametrams rasti naudosime tokią funkciją:

```
a=1 #a ir t – pradines parametru reiksmes
t=1000
lnmle=function(xx) {
   k=0.372238898
   x=sort(xx)
   n=length(x)
   atemp=0
   ttemp=0
   for (m in 1:(n-1)){
   tik=function(param){
      a<-param[1]
      t<-param[2]
      suma3=0
      for(i in 1:m) suma3=suma3+((log(x[i])-log(t))^2)
      ll<-n*log(a)+a*n* log(t)-n*log(1+ pnorm(k))-a*sum(log(x))- sum(log(x))-
      0.5*(a^2)*suma3/(k^2)
   }
   tikiv= maxLik(tik, start=c(aa,tt))
   atemp[m]= tikiv $estimate[1]
   ttemp[m]= tikiv $estimate[2]
   }
   ttempvector=c(rep(0,n))
   for (i in 1:(n-1))
```

```
   if(x[i]<=ttemp[i]) if(x[i+1]>=ttemp[i])
   ttempvector[i]=ttemp[i]
   tml=max(ttempvector)
   aml=atemp[which(ttempvector==tml)]
   return(tml,aml)
}
lnmle(x)
```

Gauname tokius parametrų įverčius: $\hat{\alpha}_{ML}$=0,7364984, $\hat{\theta}_{ML}$=1102,289. Apsirašyti šio skirstinio pasiskirstymą galime tokia funkcija:

```
plnpar = function(x) {
   k=0.3722389
   fik=pnorm(k)
   ifelse (x>0&x<=t,
   1/(1+pnorm(k))*pnorm((a/k)*log(x/t)+k),
   1-(1/(1+pnorm(k)))*(t/x)^a)
}
```

suskaičiuokime testų statistikas gautiems įverčiams:

```
a=0.7364984
t=1102.289
kstestas(x,plnpar)
adtestas(x,plnpar)
chitest(x,plnpar,650,950,1200,1700,3000,7000)
```

Gauname tokias statistikų reikšmes: Kolmogorovo – Smirnovo testo: 0,06831788; Anderseno – Darlingo testo: 0,1580152; chi kvadrato suderinamumo kriterijaus p reikšmė: 0,9956.

Taip pat nagrinėkime Kahadawala Cooray, Malwane M. A. Amanda straipsnyje „Modeling actuarial data with a composite lognormal – Pareto model“ aprašytu „Ad-hoc“ metodu gautus įverčius. Įverčius gauname pasinaudodami funkcija:

```
lnahe=function(xx){
   k=0.3722389
   x=sort(xx)
   n=length(x)
   fik=pnorm(k)
   p=fik/(1+fik)
   m=trunc((n+1)*p)
   h=(n+1)*p-m
   tad=h*x[m]+(1-h)*x[m+1]
   sum1=0
   for(i in 1:n) sum1=sum1 + log(x[i]/tad)
```

```
   sum2=0
   for(i in 1:m) sum2=sum2 + (log(x[i]/tad))^2
   aad=(sqrt(k^4*sum1^2+4*n*k^2*sum2)-k^2*sum1)/(2*sum2)
   return(aad,tad)
}
```

Gauname tokius parametrų įverčius: $\hat{\alpha}_{ad}$ =0,7742606, $\hat{\theta}_{ad}$ =979,0273. Suskaičiuokime testų statistikas gautiems įverčiams:

```
a=0.7742606
t=979.0273
kstestas(x,plnpar)
adtestas(x,plnpar)
chitest(x,plnpar,650,950,1200,1700,3000,7000)
```

Gauname tokias statistikų reikšmes: Kolmogorovo – Smirnovo testo: 0,1225273; Anderseno – Darlingo testo: 0,5561077; chi kvadrato suderinamumo kriterijaus p reikšmė: 0,9688.

- Nagrinėkime lognormalųjį skirstinį. Jo didžiausio tikėtinumo parametrus rasime naudodamiesi „fitdistr“ funkcija iš paketo „MASS“:

```
fitdistr(x, "lognormal")
```

Gauti įverčiai : $\hat{\mu}$ = 7,6195839; $\hat{\sigma}$ = 1,3709696. Apsirašykime pasiskirstymo funkciją ir suskaičiuokime testų statistikas gautiems įverčiams:

```
pln=function(x){pnorm((log(x)-m)/s)}
m= 7.6204151
s= 1.3709314
kstestas(x,pln)
adtestass(x,pln)
chitest(x,pln,650,950,1200,1700,3000,7000)
```

Gauname statistikų reikšmes: Kolmogorovo – Smirnovo testo: 0,1347996; Anderseno – Darlingo testo: 1,793444; chi kvadrato suderinamumo kriterijaus p reikšmė: 0,0344.

- Nagrinėkime Pareto skirstinį. Didžiausio tikėtinumo parametrus rasime naudodamiesi tokia funkcija:

```
parmle=function(xx){
   x=sort(xx)
   tml=min(x)
   sum=0
   n=length(x)
   for(i in 1:n) sum=sum+(log(x[i])-log(tml))
   aml=n/sum
```

```
   return(aml,tml)
}
parmle(x)
```

Gauti įverčiai : $\hat{\alpha}$ =0,5271954; $\hat{\theta}$=306. Apsirašykime pasiskirstymo funkciją ir suskaičiuokime testų statistikas gautiems įverčiams:

```
ppar=function(x){1-(t/x)^a}
a=0.5271954
t=306
kstestass(x,ppar)
chitest(x,ppar,650,950,1200,1700,3000,7000)
```

Gauname statistikų reikšmes: Kolmogorovo – Smirnovo testo: 0,2083834; chi kvadrato suderinamumo kriterijaus p reikšmė: 0,0344. Pareto skirstiniui Anderseno – Darlingo testo taikyti negalima, nes gautas įvertis $\hat{\theta}$ = 306, o pasiskirstymo funkcija taške x = 306 lygi 0, taigi negalima skaičiuoti jos logaritmo, reikalingo A - D statistikos testui.

- Nagrinėkime gama skirstinį. Jo didžiausio tikėtinumo parametrus rasime užsirašę tankio funkciją bei pasinaudoję „fitdistr“:

```
dg=function(x,a,t) {1/(x*gamma(a))*(x/t)^a*exp(-x/t)}
fitdistr(x,dg,list(a=1,t=1000),lower=0.01)
```

Gauti įverčiai: $\hat{\alpha}$ =0,3637933; $\hat{\theta}$=35185,4. Apsirašykime pasiskirstymo funkciją ir suskaičiuokime testų statistikas gautiems įverčiams:

```
fgg=function(x) {1/(x*gamma(a))*(x/t)^a*exp(-x/t)}
pg=function(x){integrate(fgg,0,x)$value}
a=0.3637933
t=35185.4
kstestas(x,pg)
adtestas(x,pg)
chitest(x,pg,650,950,1200,1700,3000,7000)
```

Gauname statistikų reikšmes: Kolmogorovo – Smirnovo testo: 0,3118727; Anderseno – Darlingo testo: 7.91738; chi kvadrato suderinamumo kriterijaus p reikšmė: 0,0000.

- Nagrinėkime Weibull skirstinį. Didžiausio tikėtinumo parametrus rasime pasinaudoję „fitdistr“:

```
fitdistr(x, "weibull")
```

Gauti įverčiai: $\hat{\alpha}$ =0.5270559, $\hat{\theta}$=3946.277. Apsirašykime pasiskirstymo funkciją ir suskaičiuokime testų statistikas gautiems įverčiams:

```
pweib=function(x){1-exp(-(x/t)^a)}
```

```
a=0.5270559
t=3946.277
kstestass(x,pweib)
adtestass(x,pweib)
chitest(x,pweib,650,950,1200,1700,3000,7000)
```

Gauname statistikų reikšmes: Kolmogorovo – Smirnovo testo: 0,2335909; Anderseno – Darlingo testo: 4,408773; chi kvadrato suderinamumo kriterijaus p reikšmė: 0,0000.

3.1.2 Analizė

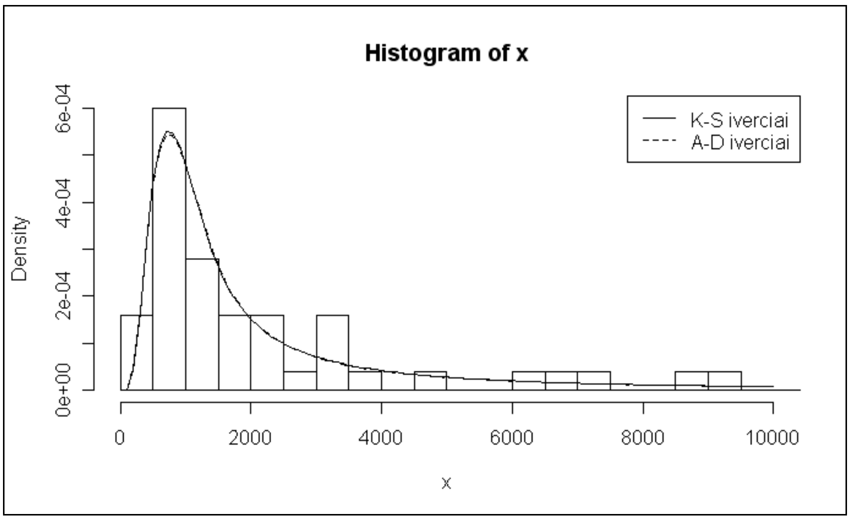
Pateiksime apibendrintus skaičiavimų rezultatus antroje lentelėje:

Skirstinys	**Parametrų įverčiai**	**K – S testo statistika**	**A – D testo statistika**	**Chi kvadrato suderinamumo kriterijaus p – reikšmė**
Lognormalusis – apibendrintas Pareto (K-S įverčiai) d = 0,4891439 1-d = 0,5108561	$\hat{\alpha}_{KS}$ =0,7327476 $\hat{\beta}_{KS}$ =1716,708 $\hat{\theta}_{KS}$ =1400,044 $\hat{\sigma}_{KS}$ =0,5766252	0,05603657	0,1298399	0,8853
Lognormalusis – apibendrintas Pareto (A-D įverčiai) d = 0,5095374 1-d = 0,4904626	$\hat{\alpha}_{KS}$ =0,7233132 $\hat{\beta}_{KS}$ =1773,036 $\hat{\theta}_{KS}$ =1473,014 $\hat{\sigma}_{KS}$ =0,5885831	0,05833153	0,1292431	0,8633
Lognormalusis – Pareto (didžiausio tikėtinumo įverčiai)	$\hat{\alpha}_{ML}$ =0,7377677, $\hat{\theta}_{ML}$ =1104,371	0,06831788	0,1575494	0,9956
Lognormalusis – Pareto (Ad-hoc įverčiai)	$\hat{\alpha}_{ad}$ = 0,7742606; $\hat{\theta}_{ad}$ =979,0273	0,1225273	0,5561077	0,9688
Lognormalusis	$\hat{\mu}$ = 7,6204151; $\hat{\sigma}$ = 1,3709314	0,1347996	1,793444	0,0344
Pareto	$\hat{\alpha}$ =0,5271954; $\hat{\theta}$=306	0,2083834	Negalima reikšmė	0,0183
Gama	$\hat{\alpha}$ =0,3637933; $\hat{\theta}$=35185,4	0,3118727	7,91738	0,0000
Weibull	$\hat{\alpha}$ = 0,5270559, $\hat{\theta}$= 3946,277	0,2335909	4,408773	0,0000

2 lentelė

Iš gautų rezultatų matome, kad sudėtinis lognormalusis – apibendrintas Pareto modelis gerai aprašo nagrinėjamą imtį. Kolmogorovo – Smirnovo bei Anderseno – Darlingo testų statistikos rodo, jog šis skirstinys yra tinkamiausias šiai imčiai. Ši išvada yra gana pagrįsta, kadangi šis skirstinys yra artimas sudėtiniam lognormaliąjam – Pareto modeliui, kuriuo ir buvo generuota ši imtis. Chi kvadrato suderinamumo kriterijaus p reikšmės rodo, kad tinkamiausias šiems duomenims aprašyti yra sudėtinis lognormalusis – Pareto modelis. Svarbu pažymėti, kad chi kvadrato suderinamumo kriterijaus reikšmės yra priklausomos nuo pasirinktų grupavimo intervalų. Vis dėlto iš visų nagrinėtų skirstinių akivaizdžiai išsiskiria du pirmieji. Kitų skirstinių testų statistikos nerodo tokių gerų rezultatų. Galima pastebėti, kad tiek Kolmogorovo – Smirnovo, tiek Anderseno – Darlingo testų statistikomis pagrįsti sudėtinio lognormaliojo – apibendrinto Pareto skirstinio įverčiai yra gana artimi, atsižvelgiant į gautus testų rezultatus. Nagrinėdami parametrų d ir $1-d$ reikšmes, galime atkreipti dėmesį į tai, jog tiek lognormaliajam, tiek apibendrintam Pareto skirstiniams priskiriamos beveik lygios svorių parametrų reikšmės. Rezultatas seka iš to, jog duomenys generuoti su mišiniu, kurio svorio parametrai jį sudarantiems skirstiniams yra fiksuoti ir lygūs.

Ketvirtame paveikslėlyje nubraižyta duomenų histograma bei darbe nagrinėjamo skirstinio tankiai su dviem įverčių rinkiniais. Galime matyti, jog formos yra gana panašios, artimos histogramai:



pav. 4

3.2. Danijos gaisrų draudimo žalų analizė

Šiame skyrelyje nagrinėsime Danijos gaisrų draudimo žalas. Šie duomenys gana gerai žinomi draudimo matematikams. Imtį sudaro 2492 draudimo žalos, užregistruotos nuo 1980 iki 1990 metų. Imtį sudaro gaisrų padaryti nuostoliai pastatams, baldams, įrengimams ir kitai privačiai nuosavybei, išreikšti Danijos kronomis. Žalų dydžiai išreikšti milijonais. Mažiausia žala yra 0,3134041 milijonų Danijos kronų, didžiausia - 263,2503660.

Kadangi šiame skyrelyje naudojamos funkcijos skirstinių didžiausio tikėtinumo įverčiams gauti, bei kitiems skaičiavimams yra tos pačios kaip ir praeitame, visų skaičiavimų neberodysime. Verta tik paminėti, jog skaičiuojant Danijos gaisrų draudimo žaloms skirstinių parametrų įverčius, funkcijose, kur reikia nurodyti pradines reikšmes, skaičiavimai atlikti pasirinkus $\alpha = 1$, $\beta = 1$, $\theta = 1$, $\sigma = 1$. Chi kvadrato suderinamumo kriterijui taikyti surikiuotos imties intervalų dešinieji galai buvo pasirinkti taip: 0,7; 0,9; 1,7; 3,6; 4; 8; ∞.

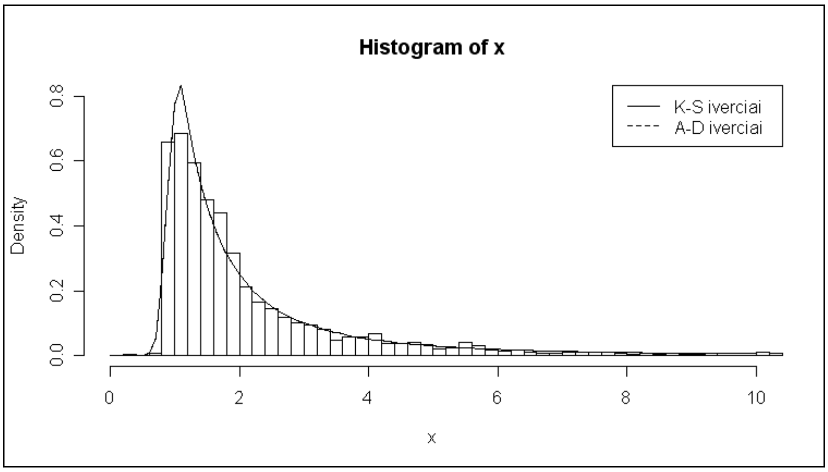
3.2.1. Analizė

Iš rezultatų, pateiktų trečioje lentelėje, akivaizdžiai matome sudėtinio lognormaliojo – apibendrinto Pareto modelio pranašumą, lyginant su kitais skirstiniais. Kolmogorovo – Smirnovo bei Anderseno – Darlingo testų statistikos lyginant skirstinius šiuo atveju yra mažiausios, o chi kvadrato suderinamumo kriterijaus p reikšmės yra gerokai didesnės, lyginant su kitais skirstiniais. Įverčiai, tiek išrinkti ieškant mažiausios Kolmogorovo – Smirnovo testo statistikos, tiek mažiausios Anderseno – Darlingo testo statistikos, yra ganėtinai artimi. Matome jog lognormaliajam skirstiniui priskiriama apie 24% skirstinio svorio, o apibendrintam Pareto – apie 76%. Be šio skirstinio, gana geri lognormaliojo – Pareto modelio analizės rezultatai, tačiau atsižvelgiant į tai kokios svorio parametrų reikšmės priskiriamos pirmajam skirstiniui, natūralu manyti, jog šis skirstinys, turintis fiksuotus svorius, nėra tinkamiausias. Likusių skirstinių neatitikimą realiems duomenims gana akivaizdžiai rodo chi kvadrato suderinamumo kriterijaus p reikšmės (jos lygios nuliui).

Skirstinys	**Parametrų įverčiai**	**K - S testo statistika**	**A - D testo statistika**	**Chi kvadrato suderinamumo kriterijaus p - reikšmė**
Lognormalusis – apibendrintas Pareto (K-S įverčiai) d = 0,2440719 1-d = 0,7559281	$\hat{\alpha}_{KS}$ = 1,554339 $\hat{\beta}_{KS}$ = 0,9639605 $\hat{\theta}_{KS}$ = 1,153572 $\hat{\sigma}_{KS}$ = 0,1854576	0,01939325	1,973170	0,6032
Lognormalusis – apibendrintas Pareto (A-D įverčiai) d = 0,2445181 1-d = 0,7554819	$\hat{\alpha}_{AD}$ = 1,552608 $\hat{\beta}_{AD}$ = 0,96222 $\hat{\theta}_{AD}$ = 1,153982 $\hat{\sigma}_{AD}$ = 0,1853665	0,01974554	1,968568	0,6179
Lognormalusis - Pareto (didžiausio tikėtinumo įverčiai)	$\hat{\alpha}_{ML}$ = 1,436332 $\hat{\theta}_{ML}$ = 1,385128	0,02873353	4,258509	0,0138
Lognormalusis - Pareto (Ad-hoc įverčiai)	$\hat{\alpha}_{ad}$ = 1,418804 $\hat{\theta}_{ad}$ = 1,402640	0,03537422	4,358444	0,0570
Lognormalusis	$\hat{\mu}$ = 0,67185368 $\hat{\sigma}$ = 0,73231667	0,1271396	85,49343	0,0000
Pareto	$\hat{\alpha}$ = 0,5458171 $\hat{\theta}$= 0,3134040	0,4082812	Negalima reikšmė	0,0000
Gama	$\hat{\alpha}$ = 1,25799429 $\hat{\theta}$= 2,43459177	0,2012952	212,447	0,0000
Weibull	$\hat{\alpha}$ = 0,94758703 $\hat{\theta}$= 2,95249589	0,2554568	begalybė	0,0000

3 lentelė

Penktame paveikslėlyje pateikta Danijos gaisrų draudimo žalų histograma, bei sudėtinio lognormaliojo – apibendrinto Pareto tankio grafikai (vėlgi su abiem nagrinėtais parametrų rinkiniais). Šiuo atveju matome jog tankių formos atkartoja histogramos formą, o tarpusavyje abu tankiai yra beveik vienodi:


Histogram of x
K-S iverciai
A-D iverciai
Density
0.0
0.2
0.4
0.6
0.8
0
2
4
6
8
10
x

pav. 5

3.3. vienos Lietuvos draudimo kompanijos ne gyvybės draudimo žalos

Galiausiai nagrinėsime vienos Lietuvos draudimo kompanijos, užsiimančios ne gyvybės draudimu, žalas. Imtį sudaro ne gyvybės draudimo žalos, kurių mažiausia yra 8,54 Lt, didžiausia – 200000 Lt. Imties tūris – 809.

Vėlgi svarbu paminėti, kad kai kurie skaičiavimai atlikti pasirinkus tokias pradines parametrų reikšmes: $\alpha = 1$, $\beta = 1$, $\theta = 1$, $\sigma = 1$. Chi kvadrato suderinamumo kriterijui taikyti surikiuotos imties intervalų dešinieji galai buvo pasirinkti taip: 200; 300; 400; 500; 700; 3000; ∞.

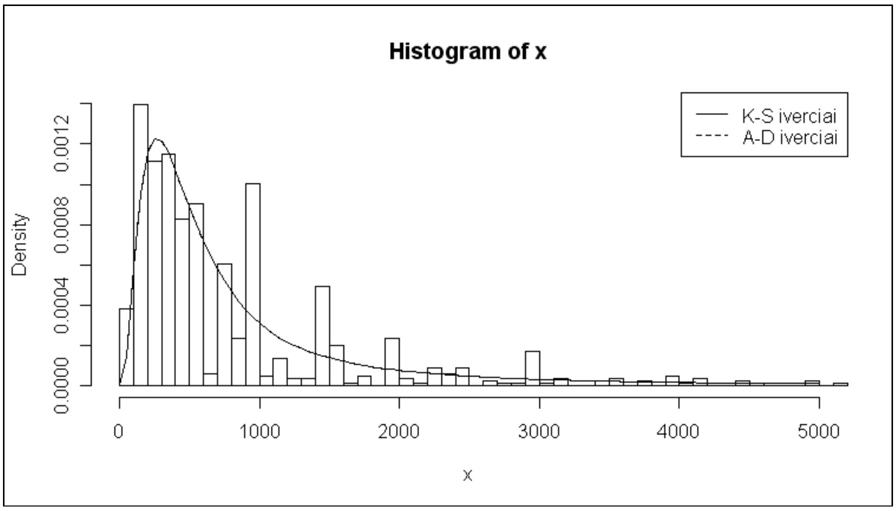
3.3.1. Analizė

Nagrinėdami ketvirtą lentelę matome, jog kaip ir praeituose skyreliuose, kuriuose analizuojami duomenys, šiame taipogi galime teigti, jog sudėtinis lognormalusis - apibendrintas Pareto skirstinys yra tinkamesnis šiems duomenims aprašyti nei kiti palyginimui pasirinkti skirstiniai, tai rodo tiek Kolmogorovo – Smirnovo bei Anderseno – Darlingo testų statistikos, kurios šiam skirstiniui yra mažiausios, tiek chi kvadrato suderinamumo kriterijaus p reikšmė, kuri tik šiuo atveju nėra lygi nuliui. Deja šio kriterijaus p reikšmės nagrinėjamiems duomenims yra mažos. Tai reiškia, jog ir skirstinys aprašo šią imtį nelabai gerai. Parametrai išrinkti abiem metodais yra tie patys. Priešingai nei nagrinėjant Danijos gaisrų draudimo žalas, šiuo atveju didesnis svoris priskiriamas lognormaliajam skirstiniui (d = 0,6578986), apibendrintam Pareto skirstiniui priskiriamas svoris lygus 0,3421014. Iš likusių skirstinių būtent lognormalusis ir yra geriausiai aprašantis šiuos duomenis. Lyginant testų kriterijų reikšmes, šiam skirstiniui nusileidžia kitiems duomenims labiau tikęs lognormalusis – Pareto modelis. Taigi galima daryti išvadą, jog sudėtinis lognormalusis - apibendrintas Pareto skirstinys gana gerai aprašo tiek duomenis, kuriuos sudaro daug mažų žalų, tiek duomenis, kuriems aprašyti tinka sunkias uodegas turintys skirstiniai.

Skirstinys	Parametrų įverčiai	K - S testo statistika	A - D testo statistika	Chi kvadrato suderinamumo kriterijaus p - reikšmė
Lognormalusis – apibendrintas Pareto (K-S įverčiai) d = 0,6578986 1- d = 0,3421014	$\hat{\alpha}_{KS}$ = 1,440740 $\hat{\beta}_{KS}$ = 780,6681 $\hat{\theta}_{KS}$ = 829,4381 $\hat{\sigma}_{KS}$ = 0,7978217	0,06557432	3,590097	0,0335
Lognormalusis – apibendrintas Pareto (A-D įverčiai) d = 0,6578986 1- d = 0,3421014	$\hat{\alpha}_{AD}$ = 1,440740 $\hat{\beta}_{AD}$ = 780,6681 $\hat{\theta}_{AD}$ = 829,4381 $\hat{\sigma}_{AD}$ = 0,7978217	0,06557432	3,590097	0,0335
Lognormalusis - Pareto (didžiausio tikėtinumo įverčiai)	$\hat{\alpha}_{ML}$ =0,727092 $\hat{\theta}_{ML}$ = 397,5667	0,1031226	12,88781	0,0000
Lognormalusis - Pareto (Ad-hoc įverčiai)	$\hat{\alpha}_{ad}$ = 0,725158 $\hat{\theta}_{ad}$ = 400	0,1004318	12,92433	0,0000
Lognormalusis	$\hat{\mu}$ = 6,4359625 $\hat{\sigma}$ = 1,04099936	0,09281261	7,268845	0,0000
Pareto	$\hat{\alpha}$ = 0,2330350 $\hat{\theta}$= 8,54	0,4685254	Negalima reikšmė	0,0000
Gama	$\hat{\alpha}$ = 0,6207303 $\hat{\theta}$= 2706,134	0,242233	35,46724	0,0000
Weibull	$\hat{\alpha}$ = 0,7005034 $\hat{\theta}$= 1024,981	0,2207134	begalybė	0,0000

4 lentelė

Šeštame paveikslėlyje pateikta ne gyvybės draudimo žalų histograma, bei sudėtinio lognormaliojo – apibendrinto Pareto tankio grafikai, kurie šiuo atveju sutampa, kadangi gauti įverčiai yra tie patys. Sudėtinga histogramos forma neleidžia šiam skirstiniui gerai atkartoti jos frmos, tačiau duomenų kitimas yra gana artimas modelio savybėms:



Histogram of x
K-S iverciai
A-D iverciai
Density
0.0000
0.0004
0.0008
0.0012
0
1000
2000
3000
4000
5000
x

pav. 6

IŠVADOS

Šiame magistro darbe, remiantis Kahadawala Cooray, Malwane M. A. Amanda straipsniu „Modeling actuarial data with a composite lognormal – Pareto model“ bei McNeil Alexander J. straipsniu „Estimating the tails of loss severity distributions using extreme value theory“, sukurtas sudėtinis lognormalusis – apibendrintas Pareto skirstinys, turintis keturis laisvus parametrus. Santykinai didelis parametrų skaičius bei gan lankstus sujungimo būdas suteikia geras modelio pritaikymo realiems draudimo duomenims galimybes. Laisvo svorio parametro d įvedimas sujungime, leidžia taikyti šį skirstinį įvairaus tipo žaloms vertinti.

Duomenų analizės rezultatai rodo, jog šis modelis, lyginant jį su kitais panašiais skirstiniais, geriausiai tinka visoms trims nagrinėtoms duomenų imtims aprašyti. Nors lognormalusis – Pareto skirstinys taip pat daugeliu atveju rodo gerus rezultatus, tačiau du papildomi parametrai leidžia geriau priartinti sudėtinio lognormaliojo – apibendrinto Pareto skirstinio tankį prie realių duomenų histogramos.

Pagrindiniu šio modelio trūkumu galima laikyti didelėms duomenų imtims gana ilgai trunkantį parametrų skaičiavimo algoritmą. Tačiau siekiant kuo geriau įvertinti draudimo žalas, sudėtinis lognormalusis – apibendrintas Pareto skirstinys turi labai geras pritaikymo galimybes.

LITERATŪRA IR ŠALTINIAI

1. V. Čekanavičius, G. Murauskas, *Statistika ir jos taikymai. I*, Vilnius, TEV, 2000.

2. Kahadawala Cooray, Malwane M. A. Amanda, Modeling actuarial data with a composite lognormal – Pareto model, *Scandinavian Actuarial Journal*, 2005, 5, p. 321-334

3. J. Kruopis, *Matematinė statistika*, Vilnius, Mokslo ir enciklopedijų leidykla, 1993.

4. McNeil Alexander J., *Estimating the tails of loss severity distributions using extreme value theory*. *ASTIN Bulletin*, 1997, 27, 117-137.