VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistro darbas

NUPJAUTO LOGNORMALIOJO IR PARETO SKIRSTINIŲ MIŠINIO KAI KURIOS SAVYBĖS

SOME PROPERTIES OF THE TRUNCATED LOGNORMAL AND THE PARETO DISTRIBUTIONS MIXTURE

Viktorija Žuklijaitė

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

Draudimo matematikoje modeliuojant žalas dažnai naudojami dviejų parametrų lognormalusis ir Pareto skirstiniai. Lognormalusis skirstinys taikomas mažoms žaloms su dideliu dažniu aprašyti, o Pareto – didelėms su mažu. Siekiant, kad skirstinys vienodai gerai aprašytu visų tipų žalas, sudaromas nupjauto lognormaliojo ir Pareto skirstinių mišinys su trimis laisvais parametrais. Šis darbas parašytas remiantis Kahadawala Cooray ir Malwane M. A. Ananda straipsniu "Modeling actuarial data with a composite lognormal – Pareto model" ("Scandinavian Actuarial Journal", 2005, 5, 321 - 334 psl.), kuriame nagrinėjamas sudėtinis lognormalusis – Pareto skirstinys.

Darbe nagrinėjami lognormalusis ir Pareto skirstiniai, aptariama, kodėl jie turėtų būti naudojami kartu kaip mišinys, pateikiamas nupjauto lognormaliojo ir Pareto skirstinių mišinio tankio funkcijos išvedimas, grafiškai iliustruojamas jos kitimas, priklausomai nuo parametrų parinkimo. Praktinėje dalyje nagrinėjamos trys imtys: sudėtiniu lognormaliuoju – Pareto skirstiniu modeliuoti duomenys, vienos Lietuvos draudimo bendrovės draudimo nuo nelaimingų atsitikimų žalos ir Danijos gaisrų draudimo žalos. Didžiausio tikėtinumo metodu skaitiškai įvertinami mišinio parametrai kiekvienai imčiai, gauti rezultatai lyginami su sudėtinio lognormaliojo – Pareto, lognormaliojo, Pareto, Gama, Weibull skirstinių didžiausio tikėtinumo įverčiais. Palyginimui naudojami Kolmogorovo – Smirnovo, Andersono – Darlingo kriterijai ir chi kvadrato suderinamumo testo p – reikšmė. Gauti rezultatai rodo, kad visais atvejais geriausiai tinka nupjauto lognormaliojo ir Pareto skirstinių mišinys draudimo žaloms analizuoti ir modeliuoti.

Resume

In insurance mathematics are often used the lognormal ant the Pareto distributions with two parameters to model loss data. The lognormal distribution is used to model small data with higher frequencies, while the Pareto distribution is used to model large data with low frequencies. In order to achieve both of these losses in one model, the truncated lognormal and the Pareto distributions mixture with three parameter is presented. This work is written with reference to Kahadawala Cooray ir Malwane M. A. Ananda article "Modeling actuarial data with a composite lognormal – Pareto model" ("Scandinavian Actuarial Journal", 2005, 5, 321 - 334 pages), where composite lognormal – Pareto model is researched.

In this work the necessity of the lognormal and the Pareto distributions mixture is discussed, the derivation of the truncated lognormal and the Pareto distributions mixture model are presented and behaviour of density function in dependent of parameter variation is discussed by illustrating. For practical application three data sets are chosen: simulated from the composite lognormal – Pareto distribution, personal accident insurance loss of one Lithuania insurance company, Danish fire loss data. For the each of data sets the parameters are estimated using maximum likelihood function, resulted estimators are compared with maximum likelihood estimators of composite lognormal – Pareto, lognormal, Pareto, Gamma and Weibull distributions. In order to compare the models the following criterions are used: Kolmogorov – Smirnov, Anderson – Darling criterions and the p – value from the chi square goodness of it test. The result indicate that the truncated lognormal and the Pareto distributions mixture fits most for every data set of insurance loss.

Įvadas

Šis darbas paremtas Kahadawala Cooray ir Malwane M. A. Ananda straipnyje "Modeling actuarial data with a composite lognormal – Pareto model" ("Scandinavian Actuarial Journal", 2005, 5, 321 - 334 psl.) nagrinėtu sudėtinio lognormaliojo – Pareto skirstinio modeliu. Remdamiesi šiuo straipsniu, 2006 metais kartu su Giedre Tamulyje ir Gediminu Kuodžiu parašėme bakalauro darbą „Žalų modeliavimas naudojant sudėtinį lognormalųjį – Pareto skirstinį“. Magistro darbe nagrinėjamas nupjauto lognormaliojo ir Pareto skirstinių mišinys.

Aktuarinėje matematikoje bei draudimo bendrovių kasdieniniame darbe dažnai naudojami lognormalusis ir Pareto skirstiniai draudimo žalų duomenims generuoti ir analizuoti. Šie duomenys dažniausiai yra labai asimetriški ir stipriai nukrypę į teigiamą puse. Didelio dažnio mažas žalas gerai aprašo lognormalusis skirstinys, o perdraudimo ir dideles žalas su mažesniu dažniu geriau modeliuoti Pareto skirstiniu, kuris turi tankesnes ir ilgesnes viršutines uodegas. Lognormalusis skirstinys apima ir didesnes žalas su mažu dažniu, tačiau šio skirstinio tankio funkcijos grafikas greičiau artėja į 0, nei Pareto, t.y. lognormaliojo skirstinio uodegos yra kur kas lengvesnės. Taikant lognormalųjį skirstinį didelėms draudimo žaloms, jos dažnai yra nepakankamai įvertinamos ir nuostoliai viršija įverčius. Dėl šios priežasties, didelės žalos modeliuojamos Pareto skirstiniu, o mažoms žaloms taikomas lognormalusis skirstinys. Stengiantis, kad skirstinys atitiktų įvairias žalas, tiek mažas, tiek dideles, nagrinėkime nupjauto lognormaliojo ir Pareto skirstinių mišinį.

Skirstinių mišinį nuo 0 iki θ sudaro nupjautas taške θ lognormalusis skirstinys, kuris turi 2 nežinomus parametrus μ ir σ. Tuo tarpu, nuo θ iki begalybės mišinyje yra Pareto skirstinys, kuris turi taip pat 2 nežinomus parametrus - α ir θ. Sudarius šių dviejų skirstinių mišinį, atsiranda dar vienas kintamasis a – mišinio svorio konstanta. Siekiant, kad mišinys skirstinių sujungimo taške neturėtų trūkio ir būtų glodus, iškeliamos diferencijuojamumo ir tolydumo sąlygos jungimo taškui. Išpildžius šias sąlygas, laisvų kintamųjų skaičius sumažėja iki trijų – α, σ ir θ. Šio mišinio forma panaši į lognormaliojo skirstinio, tačiau uodegos yra sunkesnės ir panašesnės į Pareto, tačiau lengvesnės.

Šį darbą sudaro 2 skyriai. Pirmajame skyriuje išvedamas nupjauto lognormaliojo ir Pareto skirstinių mišinio modelis, tankio funkcijos grafikas lyginamas su lognormaliuojo ir Pareto skirstinių tankio funkcijų grafikais, esant tokioms pačioms parametrų reikšmėms. Grafiškai analizuojama, kaip fiksavus 2 laisvus parametrams ir keičiant trečiąjį, kinta tankio funkcijos grafikas. Taip pat analiziškai išvedamos pasiskirstymo bei t-tojo momento funkcijos. Antrajame skyriuje išvedama didžiausio tikėtinumo funkcija, kuri vėliau naudojama vertinant nežinomus parametrus skaitiškai ir analizuojamos trys duomenų imtys – sudėtiniu lognormaliuoju – Pareto skirstiniu modeliuoti duomenys, Lietuvos draudimo kampanijos draudimo nuo nelaimingų atsitikimų žalos, įvykusios per metus, bei Danijos gaisrų draudimo žalos, užregistruotos per 10 metų. Taip pat antrajame skyriuje nupjauto lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo parametrų įverčiai lyginami su kitų skirstinių didžiausio tikėtinumo įverčiais, taikant įvairių testų statistikas.

1. Nupjauto lognormaliojo ir Pareto skirstinių mišinys

1.1. Tankio funkcijos išvedimas

Nagrinėsime nupjautojo lognormaliojo ir Pareto skirstinių mišinį. Turim atsitiktinį dydį X_1, kuris pasiskirstęs pagal nupjautą taške θ lognormalųjį skirstinį, intervale nuo 0 iki θ. Šio skirstinio tankio funkcija

$$f_1(x) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\}}{\sqrt{2\pi}x\sigma\Phi\left(\frac{\ln\theta-\mu}{\sigma}\right)}, \quad 0 < x \le \theta,$$

kur $\sigma > 0$ ir μ yra nežinomi lognormaliojo skirstinio parametrai, o $\theta > 0$ - Pareto skirstinio nežinomas parametras. Intervale nuo θ iki begalybės, atsitiktinis dydis X_2 yra pasiskirstęs pagal Pareto skirstinį, su tankio funkcija

$$f_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}, \quad x > \theta.$$

Čia $\theta > 0$ ir $\alpha > 0$ yra nežinomi Pareto skirstinio parametrai.

Tarkim, nupjauto lognormalio skirstinio mišinyje bus *a* dalis, o Pareto skirstinio – $1-a$. Tuomet, skirstinių mišinio tankis yra

$$f(x) = \begin{cases} a\, f_1(x), & 0 < x \le \theta, \\ (1-a)\, f_2(x), & x > \theta. \end{cases}$$

Reikalaukime, jog sujungimo taške θ tankio funkcijos yra lygios, tolydžios ir diferencijuojamos, t.y. sujungimas yra be trūkio ir pakankamai glodus. Tada

$$\begin{cases} a\, f_1(\theta) = (1-a)\, f_2(\theta), \\ a\, f_1'(\theta) = (1-a)\, f_2'(\theta); \end{cases}$$

kur $f'(\theta)$ yra funkcijos $f(x)$ išvestinė taške θ.

Turime dviejų lygčių sistemą su penkiais nežinomaisiais – po 2 nežinomus parametrus nupjautojo lognormaliojo ir Pareto skirstinių bei mišinio svorio konstanta *a*. Ši konstanta turi tenkinti sąlygą: $0 \le a \le 1$. Išsprendus lygčių sistemą, nežinomų parametrų skaičius sumažėja iki trijų.

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

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

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

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

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

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

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

$$\begin{cases} a\, f_1(\theta) = (1-a)\, f_2(\theta), \\ a\, f_1'(\theta) = (1-a)\, f_2'(\theta); \end{cases}$$

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

Pirmos lygties išraišką statome į antrąją ir gauname:

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

$$\frac{\ln\theta - \mu}{\sigma^2} = \alpha;$$

$$\mu = \ln\theta - \alpha\sigma^2.$$

Gautąją μ išraišką, statome į pirmąją lygtį:

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

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

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

$$a = \frac{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\} + \sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}.$$

Parametras μ kinta tarp $-\infty$ ir ∞. Kai $\theta \to 0$ ir $\alpha\sigma^2 \to \infty$, tai $\mu \to -\infty$. Kai $\theta \to \infty$, tai $\mu \to \infty$. Mišinio svorio konstanta a kinta tarp 0 ir 1. Kai $\alpha\sigma \to \infty$, tai $a \to 1$. Kai $\alpha\sigma \to 0$, tai $a \to 0$.

Statome $\mu = \ln\theta - \alpha\sigma^2$ ir $a = \dfrac{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\} + \sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}$ išraiškas į $f_1(x)$ ir $f_2(x)$ tankio funkcijas:

$$f_1(x) = \frac{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\} + \sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)} \frac{1}{\sqrt{2\pi}x\sigma\Phi(\alpha\sigma)} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \ln\theta + \alpha\sigma^2}{\sigma}\right)^2\right\} =$$

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

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

$$f_2(x)=\left(1-\frac{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}\right)\frac{\alpha\theta^\alpha}{x^{\alpha+1}}=\frac{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}\frac{\alpha\theta^\alpha}{x^{\alpha+1}}}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}. \quad (2)$$

Taigi, gauname nupjauto lognormaliojo ir Pareto skirstinių mišinio tankio funkciją su trimis nežinomais parametrais $\theta > 0$, $\alpha > 0$ bei $\sigma > 0$:

$$f(x)=\begin{cases}\dfrac{\frac{\alpha\theta^\alpha}{x^{\alpha+1}}\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{x}{\theta}\right\}}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}, & 0 < x \le \theta;\\[2ex] \dfrac{\frac{\alpha\theta^\alpha}{x^{\alpha+1}}\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}, & x > \theta.\end{cases} \quad (3)$$

1.2. Tankio funkcijos grafikas

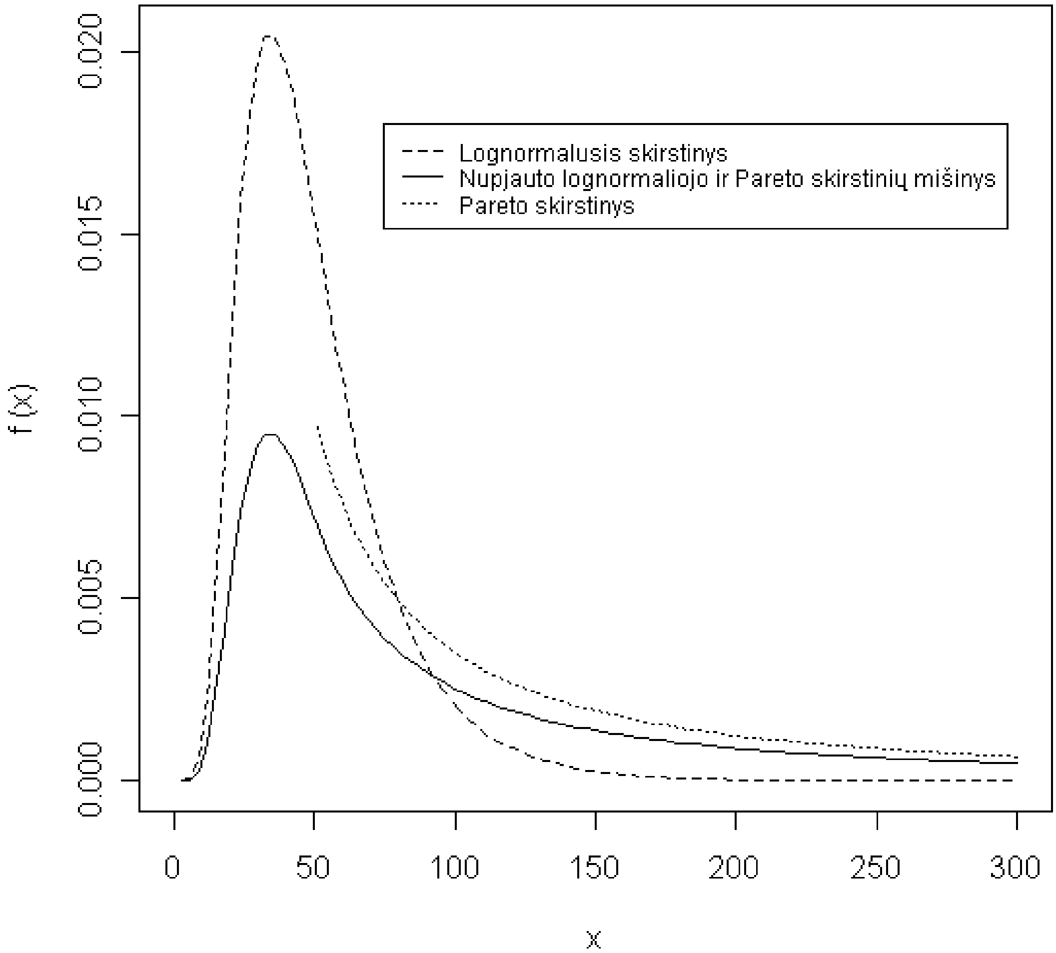
Nupjauto lognormaliojo ir Pareto skirstinių mišinio tankio funkcija priklauso nuo trijų parametrų. Su tam tikru nežinomų parametrų rinkiniu, nupjauto lognormaliojo ir Pareto skirstinių mišinio tankio funkcija neartėja į 0 taip greitai, kaip lognormaliojo. 1 pav. matome, kaip atrodo lognormaliojo, Pareto bei nupjauto lognormaliojo ir Pareto skirstinių mišinio tankių grafikai esant parametrams α=0.5, σ=0.5 ir θ=50. Esant šiam parametrų rinkiniui, mišinio svorio konstanta *a* yra 0.28. Mišinio uodega lėčiau artėja į nulį, nei lognormaliojo skirstinio tankio funkcijos, todėl geriau aprašo dideles žalas.

R pakete šį grafiką brėžiame, žymėdami nežinomus parametrus α raide a, σ - s, o θ - t (visuose kituose R paketo koduose išlaikysime tą patį žymėjimą):

```
a=0.5
s=0.5
t=50
dlnormalpar=function(x)
        {ifelse (x>0&x<=t,
        (a*(t^a)*(x^(-a-1))*exp(-(a^2)*(s^2)/2)*exp(-(log(x/t)^2)/
        (2*(s^2)))/(exp(-(a^2)*(s^2)/2)+sqrt(2*pi)*a*s* pnorm(a*s))),
        (a*(t^a)*(x^(-a-1))*exp(-(a^2)*(s^2)/2)/( exp(-
        (a^2)*(s^2)/2)+sqrt(2*pi)*a*s* pnorm(a*s))))
        }
```

```
dpareto=function(x) {ifelse(x>t, (a*(t^a))/(x^(a+1)),NA)}
m=log(t)-a*(s^2)
f=function(x) {exp(-0.5*(s^-2)*((log(x)-m)^2))/(sqrt(2*pi)*x*s)}
curve(f,0,300,lty=2)
curve(dlnormalpar,0,300,add=TRUE,lty=1)
curve(dpareto,0,300,add=TRUE,lty=3)
legend(75,0.018,c("Lognormalusis skirstinys","Nupjauto lognormaliojo
ir Pareto skirstinių mišinys","Pareto skirstinys"),lty=c(2,1,3),
cex=0.75)
```

1pav.



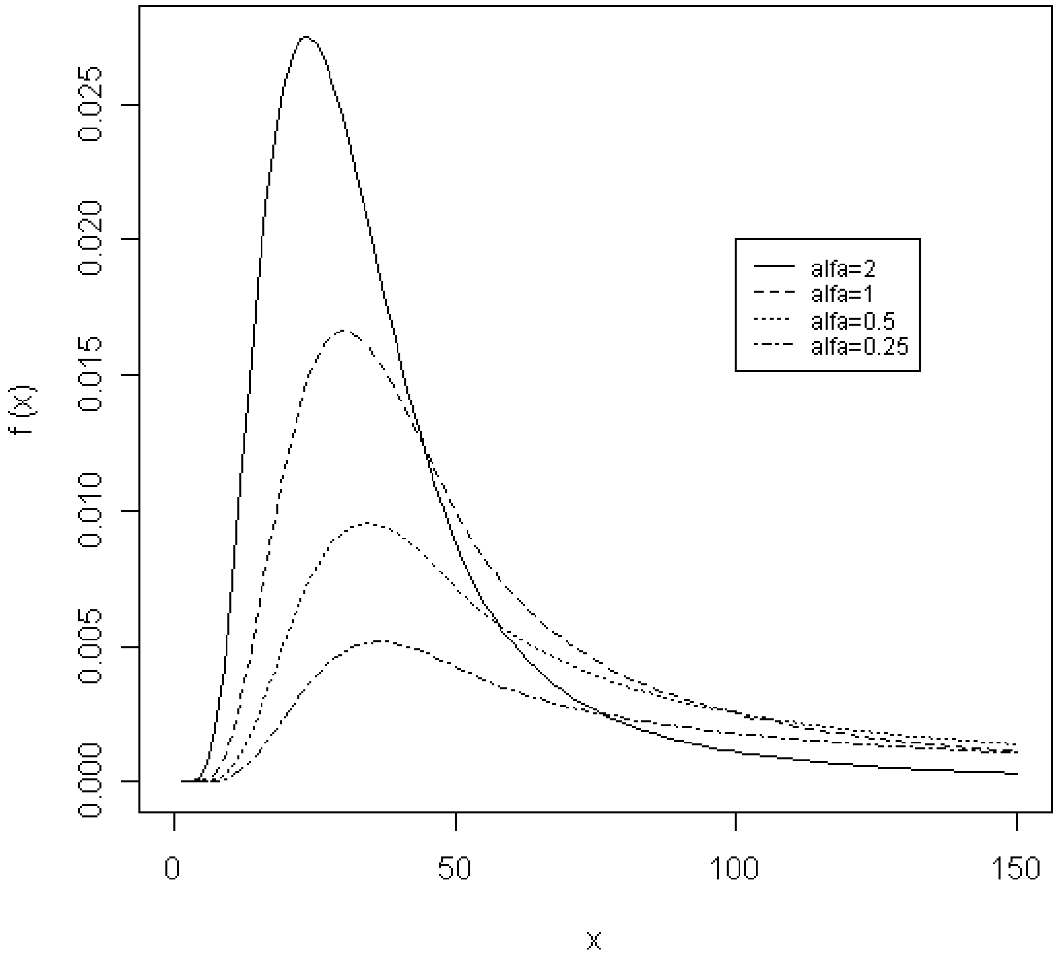
Nupjauto lognormaliojo ir Pareto skirstinių mišinio tankio funkcija priklauso nuo trijų parametrų. Fiksavus du iš jų ir keičiant trečiąjį, atitinkamai keičiasi ir tankio funkcijos grafikas.

2 pav. matyti, kaip kinta tankio funkcijos grafikas, keičiant parametrą α, esant fiksuotiems σ=0.5 ir θ=50. Mažėjant α, grafiko aukštis taip pat mažėja, viršūnė nežymiai slenka į dešinę pusę bei darosi mažiau statmena. Mišinio uodega lėčiau artėja link 0. Svorio konstanta *a*, mažėjant α, taip pat mažėja (1 lentelė).

1 lentelė

α	a
2	0.7766387
1	0.4954614
0.5	0.2790679
0.25	0.1479207

2 pav.



Statistiniame pakete šį grafiką brėžiame:

```
f=function(x)
        {ifelse (x>0&x<=t,
        (a*(t^a)*(x^(-a-1))*exp(-(a^2)*(s^2)/2)*exp(-(log(x/t)^2)/
        (2*(s^2)))/(exp(-(a^2)*(s^2)/2)+sqrt(2*pi)*a*s* pnorm(a*s))),
        (a*(t^a)*(x^(-a-1))*exp(-(a^2)*(s^2)/2)/( exp(-
        (a^2)*(s^2)/2)+sqrt(2*pi)*a*s* pnorm(a*s))))
        }
s=0.5
t=50
```

```
a=2
curve(f,0,150)
a=1
curve(f,0,150,add=TRUE,lty=2)
a=0.5
curve(f,0,150,add=TRUE,lty=3)
a=0.25
curve(f,0,150,add=TRUE,lty=4)
legend(100,0.02,c("alfa=2","alfa=1","alfa=0.5","alfa=0.25"),lty=c(1,2,
3,4), cex=0.75)
```

3 pav. vaizduoja, kaip kinta mišinio tankio funkcijos grafikas, esant fiksuotiems α =0.5 ir θ=50, o kintant σ. Didėjant σ, grafiko viršūnė taip pat didėja ir spaudžiasi prie ordinačių ašies. Mišinio svorio konstanta a, mažėjant σ, irgi mažėja (2 lentelė).

2 lentelė

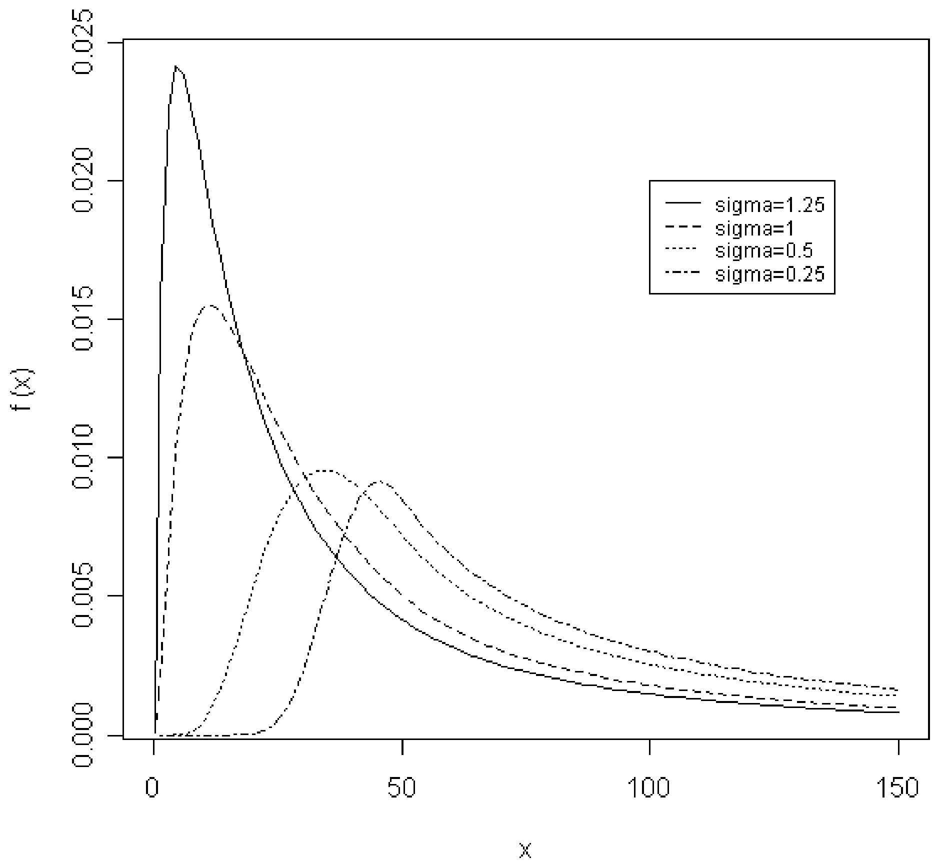
σ	a
1.25	0.5829805
1	0.4954614
0.5	0.2790679
0.25	0.1479207

Su R galima aprašyti taip:

```
f=function(x)
         {ifelse (x>0&x<=t,
         (a*(t^a)*(x^(-a-1))*exp(-(a^2)*(s^2)/2)*exp(-
         (log(x/t)^2)/(2*(s^2))))/( exp(-(a^2)*(s^2)/2)+sqrt(2*pi)*a*s*
         pnorm(a*s))),
         (a*(t^a)*(x^(-a-1))*exp(-(a^2)*(s^2)/2)/( exp(-
         (a^2)*(s^2)/2)+sqrt(2*pi)*a*s* pnorm(a*s))))
         }
a=0.5
t=50
s=1.25
curve(f,0,150)
s=1
curve(f,0,150,add=TRUE,lty=2)
s=0.5
curve(f,0,150,add=TRUE,lty=3)
s=.25
curve(f,0,150,add=TRUE,lty=4)
```

```
legend(100,0.02,c("sigma=1.25","sigma=1","sigma=0.5","sigma=0.25"),lty
=c(1,2,3,4), cex=0.75)
```

3 pav.



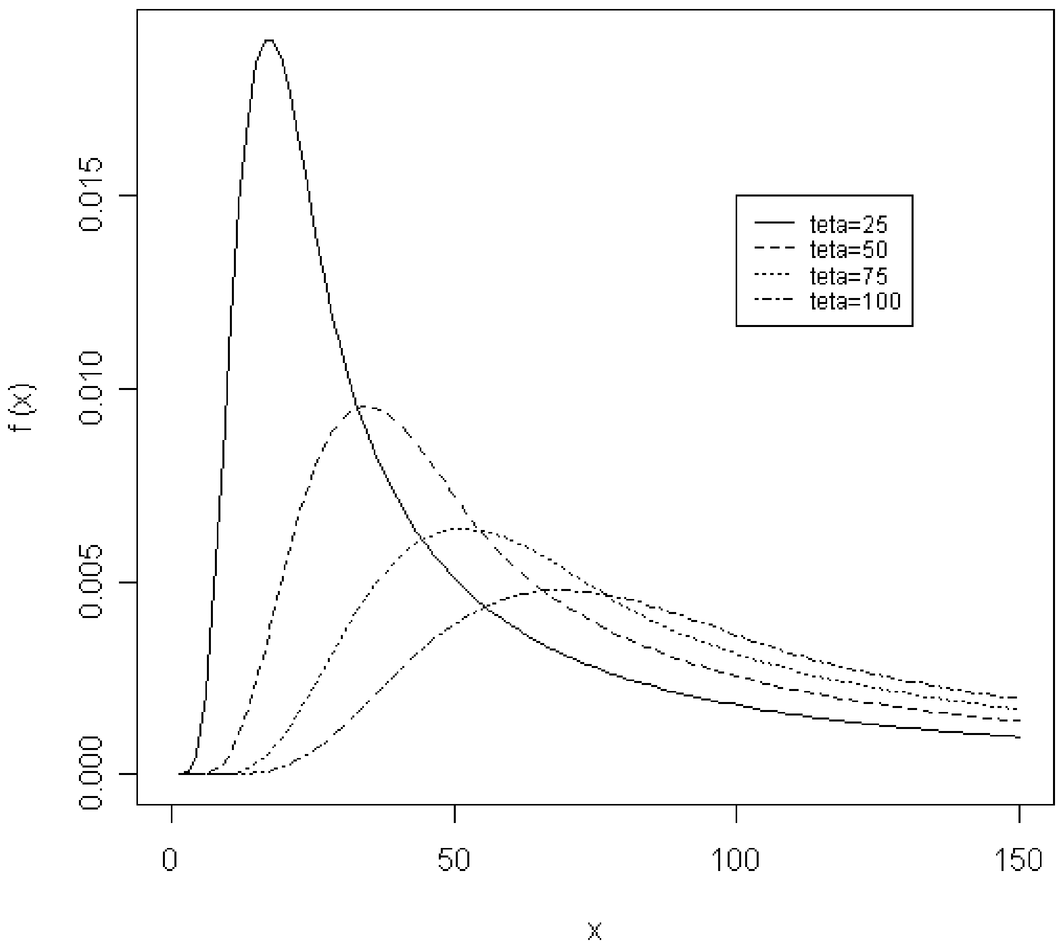
4 pav. matyti, kaip kintant θ, esant fiksuotiems α =0.5 ir σ =0.5, kinta mišinio tankio funkcijos grafikas. Didėjant θ, grafiko viršūnė mažėja, glodėja ir slenkasi į dešinę pusę. Kadangi mišinio svorio konstanta a nepriklauso nuo θ, tai ji išlieka pastovi visoms θ reikšmėms.

R pakete šį grafiką brėžiame:

```
f=function(x)
         {ifelse (x>0&x<=t,
         (a*(t^a)*(x^(-a-1))*exp(-(a^2)*(s^2)/2)*exp(-(log(x/t)^2)/
         (2*(s^2)))/(exp(-(a^2)*(s^2)/2)+sqrt(2*pi)*a*s* pnorm(a*s))),
         (a*(t^a)*(x^(-a-1))*exp(-(a^2)*(s^2)/2)/( exp(-
         (a^2)*(s^2)/2)+sqrt(2*pi)*a*s* pnorm(a*s))))
         }
a=0.5
s=0.5
t=25
curve(f,0,150)
t=50
```

```
curve(f,0,150,add=TRUE,lty=2)
t=75
curve(f,0,150,add=TRUE,lty=3)
t=100
curve(f,0,150,add=TRUE,lty=4)
legend(100,0.015,c("teta=25","teta=50","teta=75","teta=100"),lty=c(1,2
,3,4), cex=0.75)
```

4 pav.



1.3. Pasiskirstymo funkcija

Nupjauto lognormaliojo ir Pareto skirstinių mišinio pasiskirstymo funkciją *F(x)* rasime integruodami tankio funkciją *f(x)* nuo $-\infty$ iki x:

$$F(x)=\int_{-\infty}^{x} f(s)ds=\begin{cases}\int_0^x f_1(s)ds, & 0<x\le\theta,\\ \int_0^\theta f_1(s)ds+\int_\theta^x f_2(s)ds, & x>\theta,\end{cases}$$

čia funkcijos $f_1(s)$ ir $f_2(s)$ atitinkamai apibrėžtos formulėmis (1) ir (2).

Tuomet,

$$\int_0^x f_1(s)ds=\frac{\alpha\theta^\alpha\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}\int_0^x\frac{1}{s^{\alpha+1}}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{s}{\theta}\right\}ds.$$

Pažymėkime

$$C:=\frac{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}.$$

Tada,

$$\int_0^x f_1(s)ds=C\alpha\theta^\alpha\int_0^x\frac{1}{s^{\alpha+1}}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{s}{\theta}\right\}ds=$$

$$=C\alpha\int_0^x\frac{1}{s}\left(\frac{\theta}{s}\right)^\alpha\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{s}{\theta}\right\}ds=C\alpha\int_0^x\frac{1}{s}\exp\left\{\ln\left(\frac{s}{\theta}\right)^{-\alpha}\right\}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{s}{\theta}\right\}ds=$$

$$=C\alpha\int_0^x\frac{1}{s}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{s}{\theta}-\alpha\ln\frac{s}{\theta}-\frac{\alpha^2\sigma^2}{2}+\frac{\alpha^2\sigma^2}{2}\right\}ds=$$

$$=C\alpha\exp\left\{\frac{\alpha^2\sigma^2}{2}\right\}\int_0^x\frac{1}{s}\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma}\ln\frac{s}{\theta}+\alpha\sigma\right)^2\right\}ds.$$

Darome pakeitimą $t=\frac{1}{\sigma}\ln\frac{s}{\theta}+\alpha\sigma$, tuomet

$$\int_0^x f_1(s)ds=C\alpha\exp\left\{\frac{\alpha^2\sigma^2}{2}\right\}\int_{-\infty}^{\frac{1}{\sigma}\ln\frac{x}{\theta}+\alpha\sigma}\frac{1}{\theta\exp\{\sigma(t-\alpha\sigma)\}}\exp\left\{-\frac{t^2}{2}\right\}\theta\exp\{\sigma(t-\alpha\sigma)\}\sigma\,dt=$$

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

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

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

$$\int_0^\theta f_1(s)ds+\int_\theta^x f_2(s)ds=\frac{\alpha\theta^\alpha\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}\left(\int_0^\theta\frac{1}{s^{\alpha+1}}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{s}{\theta}\right\}ds+\int_\theta^x\frac{1}{s^{\alpha+1}}ds\right)=$$

$$=C\alpha\theta^\alpha\left(\int_0^\theta\frac{1}{s^{\alpha+1}}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{s}{\theta}\right\}ds+\int_\theta^x\frac{1}{s^{\alpha+1}}ds\right);$$

$$\int_0^\theta f_1(s)ds=C\alpha\theta^\alpha\int_0^\theta\frac{1}{s^{\alpha+1}}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{s}{\theta}\right\}ds=C\alpha\exp\left\{\frac{\alpha^2\sigma^2}{2}\right\}\int_0^\theta\frac{1}{s}\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma}\ln\frac{s}{\theta}+\alpha\sigma\right)^2\right\}ds.$$

Darome pakeitimą $t=\frac{1}{\sigma}\ln\frac{s}{\theta}+\alpha\sigma$, tuomet

$$\int_0^\theta f_1(s)ds=C\sqrt{2\pi}\alpha\sigma\exp\left\{\frac{\alpha^2\sigma^2}{2}\right\}\int_{-\infty}^{\alpha\sigma}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}dt=C\sqrt{2\pi}\alpha\sigma\exp\left\{\frac{\alpha^2\sigma^2}{2}\right\}\Phi(\alpha\sigma);$$

$$\int_\theta^x f_2(s)ds=C\alpha\theta^\alpha\int_\theta^x\frac{1}{s^{\alpha+1}}ds=C\alpha\theta^\alpha\left(-\frac{1}{\alpha s^\alpha}\right)\Bigg|_\theta^x=C\alpha\theta^\alpha\left(-\frac{1}{\alpha x^\alpha}+\frac{1}{\alpha\theta^\alpha}\right)=$$

$$=C\theta^\alpha\left(\frac{1}{\theta^\alpha}-\frac{1}{x^\alpha}\right)=C\left(1-\left(\frac{\theta}{x}\right)^\alpha\right);$$

$$\int_0^\theta f_1(s)ds+\int_\theta^x f_2(s)ds=C\left(\sqrt{2\pi}\alpha\sigma\exp\left\{\frac{\alpha^2\sigma^2}{2}\right\}\Phi(\alpha\sigma)+\left(1-\left(\frac{\theta}{x}\right)^\alpha\right)\right)=$$

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

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

Iš čia, nupjauto lognormaliojo ir Pareto skirstinių mišinio pasiskirstymo funkcija bus:

$$F(x)=\begin{cases}\dfrac{\sqrt{2\pi}\alpha\sigma\Phi\left(\dfrac{1}{\sigma}\ln\dfrac{x}{\theta}+\alpha\sigma\right)}{\exp\left\{-\dfrac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}, & 0<x\le\theta;\\[2ex] 1-\dfrac{\exp\left\{-\dfrac{\alpha^2\sigma^2}{2}\right\}\left(\dfrac{\theta}{x}\right)^{\alpha}}{\exp\left\{-\dfrac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}, & x>\theta.\end{cases} \tag{4}$$

1.4. t-tojo momento funkcija

Rasime nupjauto lognormaliojo ir Pareto skirstinių mišinio t-tojo momento funkciją:

$$E\left(X^t\right)=\int_{-\infty}^{\infty}x^t f(x)dx=\int_0^{\theta}x^t f_1(x)dx+\int_{\theta}^{\infty}x^t f_2(x)dx=I_1(x)+I_2(x),$$

kur funkcijos $f_1(x)$ ir $f_2(x)$ atitinkamai apibrėžtos formulėmis (1) ir (2).

Pažymėkime

$$C:=\frac{\exp\left\{-\dfrac{\alpha^2\sigma^2}{2}\right\}}{\exp\left\{-\dfrac{\alpha^2\sigma^2}{2}\right\}+\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}.$$

Taigi,

$$I_1(x)=\int_0^{\theta}x^t f_1(x)dx=C\int_0^{\theta}\frac{x^t}{x^{\alpha+1}}\alpha\theta^{\alpha}\exp\left\{-\frac{1}{2\sigma^2}\ln^2\frac{x}{\theta}\right\}dx=$$

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

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

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

Darome pakeitimą $b=\dfrac{1}{\sigma}\ln\dfrac{x}{\theta}+(\alpha-t)\sigma$, tada

$$I_1(x) = C\alpha\theta^t \exp\left\{\frac{(\alpha-t)^2\sigma^2}{2}\right\}\int_{-\infty}^{(\alpha-t)\sigma}\frac{1}{\theta\exp\left\{\sigma b-(\alpha-t)\sigma^2\right\}}\exp\left\{-\frac{b^2}{2}\right\}\theta\sigma\exp\left\{\sigma b-(\alpha-t)\sigma^2\right\}db =$$

$$= C\alpha\sigma\theta^t\sqrt{2\pi}\exp\left\{\frac{(\alpha-t)^2\sigma^2}{2}\right\}\int_{-\infty}^{(\alpha-t)\sigma}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{b^2}{2}\right\}db = C\alpha\sigma\theta^t\sqrt{2\pi}\exp\left\{\frac{(\alpha-t)^2\sigma^2}{2}\right\}\Phi((\alpha-t)\sigma);$$

$$I_2(x) = \int_{\theta}^{\infty} x^t f_2(x)dx = C\int_{\theta}^{\infty}\frac{x^t}{x^{\alpha+1}}\alpha\theta^\alpha dx = C\int_{\theta}^{\infty}\alpha\theta^\alpha x^{t-\alpha-1}dx = \left.\frac{C\alpha\theta^\alpha x^{t-\alpha}}{t-\alpha}\right|_{\theta}^{\infty} =$$

$$= -\frac{C\alpha\theta^\alpha\theta^{t-\alpha}}{t-\alpha} = \frac{C\alpha\theta^t}{\alpha-t}, \quad t<\alpha.$$

Tuomet, mišinio t-tasis momentas bus:

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

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

$$= \frac{\alpha\theta^\alpha\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\}\left(\sqrt{2\pi}\sigma\exp\left\{\frac{(\alpha-t)^2\sigma^2}{2}\right\}\Phi((\alpha-t)\sigma) + \frac{1}{\alpha-t}\right)}{\exp\left\{-\frac{\alpha^2\sigma^2}{2}\right\} + \sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}, \quad t<\alpha.$$

2. Duomenų analizė

2.1. Didžiausio tikėtinumo funkcija

Duomenų analizei naudosime didžiausio tikėtinumo funkciją. Nežinomus parametrus vertinsime skaitiškai, naudodami R statistinio programos paketą `maxLik`.

Tarkim, turime sutvarkytą n dydžio imtį $X_1, X_2, ..., X_n$ (t.y. $X_1 \le X_2 \le ... \le X_n$) iš nupjauto lognormaliojo ir Pareto skirstinių mišinio modelio su trimis nežinomais parametrais, kurio tankio funkcija aprašyta formule (3). Sakysim, kad nežinomas θ yra tarp m-tojo ir (m+1)-ojo imties nario, t.y. $x_m \le \theta \le x_{m+1}$.

Tuomet, didžiausio tikėtinumo funkcija yra

$$L(\alpha,\sigma,\theta) = f(x_1,\alpha,\sigma,\theta) * ... * f(x_n,\alpha,\sigma,\theta) = \prod_{i=1}^{m} f_1(x_i) \prod_{i=m+1}^{n} f_2(x_i),$$

čia funkcijos $f_1(x)$ ir $f_2(x)$ atitinkamai apibrėžtos formulėmis (1) ir (2).

Pažymėkime $C := \dfrac{\alpha\theta^{\alpha} \exp\left\{-\dfrac{\alpha^2\sigma^2}{2}\right\}}{\exp\left\{-\dfrac{\alpha^2\sigma^2}{2}\right\} + \sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)}$.

Tuomet,

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

$$= C^n \prod_{i=1}^{n} \frac{1}{x_i^{\alpha+1}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\ln^2\frac{x_i}{\theta}\right\}.$$

Pažymėję $\Lambda = \ln L(\alpha,\sigma,\theta)$, gausime logaritminę didžiausio tikėtinumo funkciją

$$\Lambda = n\ln C + \ln\prod_{i=1}^{n}\frac{1}{x_i^{\alpha+1}} - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\ln^2\frac{x_i}{\theta} =$$

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

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

Šią funkciją naudosime toliau aprašytoje duomenų analizėje.

2.2. Duomenų analizei naudojami duomenų rinkiniai, statistiniai kriterijai ir skirstiniai

Nagrinėdami nupjauto lognormaliojo ir Pareto skirstinių mišinio modelį, analizuosime keletą duomenų rinkinių:

- Sudėtiniu lognormaliuoju – Pareto skirstiniu generuoti duomenys. Šį skirstinį aprašė Kahadawala Cooray ir Malwane M. A. Ananda straipsnyje "Modeling actuarial data with a composite lognormal – Pareto model" ("Scandinavian Actuarial Journal", 2005, 5, 321 - 334 psl.);
- Danijos gaisrų draudimo žalos;
- Vienos Lietuvos draudimo bendrovės draudimo nuo nelaimingų atsitikimų 2007 metų žalų duomenys.

Šiems duomenų rinkiniams ieškosime mišinio didžiausio tikėtinumo įverčių ir lyginsime juos su kitų skirstinių didžiausio tikėtinumo įverčiais. Palyginimui naudosime šiuos skirstinius:

- Sudėtinį lognormalųjį – Pareto su tankio funkcija

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

  Šio skirstinio parametrai $\alpha > 0$ ir $\theta > 0$, o k - konstanta lygi 0.372238898.

- Lognormalųjį su tankio funkcija

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

  Šio skirstinio parametrai $\sigma > 0$ ir $\mu > 0$.

- Pareto su tankio funkcija

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

  Šio skirstinio parametrai $\alpha > 0$ ir $\theta > 0$.

- Gama su tankio funkcija

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

Čia $\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx.$

Šio skirstinio parametrai $\alpha > 0$ ir $\theta > 0$.

- Weibull su tankio funkcija

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

Šio skirstinio parametrai $\alpha > 0$ ir $\theta > 0$.

Gautus įvairių skirstinių didžiausio tikėtinumo įverčius lyginsime taikydami kriterijus:

- Kolmogorovo – Smirnovo testo statistika. Šis kriterijus naudojamas nustatyti, kaip empirinė imtis atitinka pasirinktą teorinį skirstinį. Kuo testo statistikos reikšmė D_n yra mažesnė, tuo skirstinys geriau aprašo empirinę imtį. Testo statistika apskaičiuojama pagal formulę:

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

čia $F(x_i)$ pasiskirstymo funkcijos reikšmė taške x_i,

n – empirinės imties dydis.

- Andersono – Darlingo testo statistika. Šis kriterijus yra Kolmogorovo – Smirnovo testo atmaina, tačiau tikslesnis. Jis taip pat naudojamas patikrinti, ar empirinė imtis yra pasiskirsčiusi pagal pasirinktą teorinį skirstinį. Kuo testo statistika A^2 yra mažesnė, tuo teorinis skirstinys geriau atitinka empirinę imtį. Statistika apskaičiuojama:

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

čia $F(x_i)$ pasiskirstymo funkcijos reikšmė taške x_i,

n – empirinės imties dydis.

- Chi kvadrato suderinamumo kriterijaus p – reikšmė. Tai alternatyva Kolmogorovo – Smirnovo bei Andersono – Darlingo kriterijams. Šis testas naudojamas nesikertančiais intervalais grupuotiems duomenims. Jo reikšmė priklauso nuo grupavimo intervalų, todėl rekomenduotina, kad visuose intervaluose būtų apylygis skaičius imties narių. Kuo chi kvadrato kriterijaus p – reikšmė didesnė, tuo geresnis pasirinktas skirstinys. Testo statistika X^2 apskaičiuojama:

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

čia V_i - imties reikšmių, esančių į i-tąjame intervale, skaičius,

n – empirinės imties dydis,

p_i - tikimybė, kad imties reikšmė pateks į i-tąjį intervalą, t.y.

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

k – sutvarkytos imties intervalų skaičius,

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

p – reikšmė yra lygi tikimybei $P\{\chi^2_{k-s-1} > X^2\}$,

čia χ^2_{k-s-1} - chi kvadrato pasiskirstymo funkcija su $k-s-1$ laisvės laipsniais,

s - nežinomų parametrų skaičius.

2.3. Sudėtiniu lognormaliuoju - Pareto skirstiniu generuotų duomenų analizė

Šiame skyriuje nagrinėsime duomenis, generuotus sudėtiniu lognormaliuoju – Pareto skirstiniu, aprašytus Kahadawala Cooray ir Malwane M. A. Ananda straipsnyje "Modeling actuarial data with a composite lognormal – Pareto model" ("Scandinavian Actuarial Journal", 2005, 5, 321 - 334 psl.). Imtį sudaro 50 duomenų nuo 306 iki 400976, generuotų su parametrais α = 1 ir θ = 1000. Imtį užrašome variacine eilute:

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.

2.3.1. Didžiausio tikėtinumo įverčiai

Naudodami R statistinį paketą `maxLik`, rasime didžiausio tikėtinumo įverčius nupjauto lognormaliojo ir Pareto skirstinių mišiniui:

```
aa=1
ss=1
tt=1000
lnmle=function(x) {
    x=sort(xx)
    n=length(x)
```

```
    atemp=0
    stemp=0
    ttemp=0
    for (m in 1:(n-1)){
          tik=function(param){
              a<-param[1]
              s<-param[2]
              t<-param[3]
              if(s<= 0) return(NA)
              suma3=0
              for(i in 1:m) suma=suma+(log(x[i])-log(t))^2
              ll<-(-n)*(a^2)*(s^2)/2+n*log(a)+n*a*log(t)-n*log(exp(-
              (a^2)*(s^2)/2)+sqrt(2*pi)*a*s*pnorm(a*s))-a*sum(log(x))-
              sum(log(x))-suma/(2*(s^2))
          }
          tikiv= maxLik(tik, start=c(aa,ss,tt))
          atemp[m]= tikiv $estimate[1]
          stemp[m]= tikiv $estimate[2]
          ttemp[m]= tikiv $estimate[3]
    }
    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)]
    sml=stemp[which(ttempvector==tml)]
    return(aml,sml,tml)
}
lnmle(x)
```

Čia `aa`, `ss` ir `tt` yra pradinės vertinamų nežinomų parametrų α, σ, θ reikšmės, su kuriomis funkcija `maxLik` pradeda iteracijas, o `x` – imies vektorius,), `ll` – tikėtinumo funkcija (toliau naudosime tokius pačius pažymėjimus).

Gauname parametrų įverčius:

$\hat{\alpha}$ =0.7635727,

$\hat{\sigma}$ =0.5407144,

$\hat{\theta}$=1203.249.

Naudodami tą patį R paketą `maxLik`, įvertinsime ir sudėtinio lognormaliojo – Pareto skirstinio nežinomus parametrus α ir θ:

```
aa=1
tt=1000
lnmle=function(x) {
     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) suma=suma+((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)*suma/(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 parametrų įverčius:

$\hat{\alpha}$ =0.7377677,

$\hat{\theta}$=1104.371.

Pareto skirstinio tankio funkcija egzistuoja tik kai x > θ. Kadangi funkcija yra griežtai mažėjanti, tai pasiekia maksimumą taške x = θ. Todėl θ didžiausio tikėtinumo įvertis ir bus šiame taške. Su R funkcija Pareto skirstinio įverčiams rasti yra:

```
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 rezultatai:

$\hat{\theta}$ = 306,

$\hat{\alpha}$ = 0.5271954.

Standartinių lognormaliojo, Gama ir Weibull skirstinių parametrų įverčius sudėtiniu lognormalioju – Pareto skirstiniu generuotiems duomenims galima gauti naudojantis funkcija `fitdistr`, esančia R pakete `MASS`.

Lognormaliojo skirstinio įverčiai:

```
fitdistr(x, "lognormal", list (meanlog=0, sdlog=1))
```

Gautos reikšmės:

$\hat{\mu}$ = 7.6195839,

$\hat{\sigma}$ = 1.3709696.

Gama skirstinio įverčiai:

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

Čia `fg` yra Gama skirstinio tankio funkcija.

Gautos reikšmės:

$\hat{\theta}$ = 35185.4,

$\hat{\alpha}$ = 0.3637933.

Weibull skirstinio įverčiai:

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

Gautos reikšmės:

$\hat{\theta}$= 0.5344084,

$\hat{\tau}$ = 4433.474.

2.3.2. Kolmogorovo – Smirnovo kriterijus

Kolmogorovo – Smirnovo testo statistika bendru atveju galima suskaičiuoti naudojant šią statistinio paketo R funkciją:

```
kstestas=function(z,func){
          n=length(z)
          y=sort(z)
          funcc=func
          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, o `func` – teorinė pasiskirstymo funkcija.

Naudojant lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo įverčius, Kolmogorovo – Smirnovo testo statistika su R paketu apskaičiuojama:

```
a=0.7635727
s=0.5407144
t=1203.249
pnlpm=function(x) {
    ifelse(x>0&x<=t, a*s*sqrt(2*pi)*pnorm(log(x/t)/s+a*s)/(exp(-(a^2)
    *(s^2)/2)+ sqrt(2*pi)*a*s*pnorm(a*s)),1-(exp(-(a^2)*(s^2)/2)
    *(t/x)^a)/(exp(-(a^2)*(s^2)/2)+ sqrt(2*pi)*a*s*pnorm(a*s)))
}
kstestas(x, pnlpm)
```

Čia `a`, `s`, `t` yra atitinkamai α, σ ir θ didžiausio tikėtinumo įverčiai, o `pnlpm` - nupjauto lognormaliojo ir Pareto skirstinių mišinio pasiskirstymo funkcija (4).

Gauta reikšmė**:** 0.06271353.

Sudėtinio lognormaliojo – Pareto skirstinio įverčiams Kolmogorovo – Smirnovo testo statistika apskaičiuojama analogiškai:

```
a=0.7377677
t=1104.371
pslp=function(x) {
```

```
    k=0.372238898
    ifelse (x>0&x<=t,1/(1+pnorm(k))*pnorm((a/k)*log(x/t)+k),1-
    (1/(1+pnorm(k)))*(t/x)^a)
}
kstestas(x, pslp)
```

Gauta reikšmė**:** 0,0672262.

Čia `pslp` – sudėtinio lognormaliojo – Pareto skirstinio pasiskirstymo funkcija:

$$F(x) = \begin{cases} \dfrac{1}{(1+\Phi(k))}\Phi\big((\alpha/k)\ln(x/\theta)+k\big), & 0 < x \le \theta; \\ 1 - \dfrac{1}{(1+\Phi(k))}(\theta/x)^{\alpha}, & \theta \le x < \infty. \end{cases}$$

Pareto skirstinio didžiausio tikėtinumo įverčių Kolmogorovo – Smirnovo statistika skaičiuojama:

```
a=0.5271954
t=306
ppar=function(x){1-(t/x)^a}
kstestass(x,ppar)
```

Gauta reikšmė: 0,2083834.

Standartiniams lognormaliajam, Gama, Pareto skirstiniams Kolmogorovo – Smirnovo testo statistikoms skaičiuoti R pakete yra naudojama funkcija `ks.test`. Įvedus duomenų vektorių ir įverčius bei pasirinkus norimą skirstinį, apskaičiuojama testo statistika.

Kolmogorovo – Smirnovo kriterijus lognormaliajam skirstiniui:

```
ks.test(x,"plnorm",7.6195839,1.3709696)
```

Gauta reikšmė: 0.1346.

Kolmogorovo – Smirnovo kriterijus Gama skirstiniui:

```
ks.test(x,"pgamma",0.3668,1/34892.34)
```

Gauta reikšmė: 0.3138.

Kolmogorovo – Smirnovo kriterijus Weibull skirstiniui:

```
ks.test(x,"pweibull",0.5344084,3946.277)
```

Gauta reikšmė 0.2299.

2.3.3. Andersono – Darlingo kriterijus

Bendru atveju Andersono – Dartingo kriterijaus reikšmę galima apskaičiuoti naudojant šią R paketo funkciją:

```
adtestas=function(z,func){
```

```
    n=length(z)
    y=sort(z)
    funcc=func
    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, kaip ir Kolmogorovo – Smirnovo kriterijaus atveju, `z` – empirinė imtis, o `func` – teorinė pasiskirstymo funkcija.

Naudojant nupjauto lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo įverčius, Andersono - Darlingo testo statistika su R paketu apskaičiuojama:

```
a=0.7635727
s=0.5407144
t=1203.249
pnlpm=function(x) {
    ifelse(x>0&x<=t, a*s*sqrt(2*pi)*pnorm(log(x/t)/s+a*s)/(exp(-(a^2)
    *(s^2)/2)+ sqrt(2*pi)*a*s*pnorm(a*s)),1-(exp(-(a^2)*(s^2)/2)
    *(t/x)^a)/(exp(-(a^2)*(s^2)/2)+ sqrt(2*pi)*a*s*pnorm(a*s)))
}
adtestas(x, pnlpm)
```

Gauta reikšmė: 0.1379439.

Sudėtinio lognormaliojo – Pareto skirstinio įverčiams Andersono - Darlingo testo statistika apskaičiuojama:

```
a=0.7377677
t=1104.371
pslp=function(x) {
    k=0.372238898
    ifelse (x>0&x<=t,1/(1+pnorm(k))*pnorm((a/k)*log(x/t)+k),1-
    (1/(1+pnorm(k)))*(t/x)^a)
}
adtestas(x, pslp)
```

Gauta reikšmė: 0.1575494.

Lognormaliojo skirstinio įverčiams Andersono - Darlingo testo statistika apskaičiuojama:

```
m=7.620415
s=1.3709
pln=function(x){pnorm((log(x)-m)/s)}
```

```
adtestas(x,pln)
```

Gauta reikšmė: 1.793403.

Gama skirstinio įverčiams Andersono - Darlingo testo statistika apskaičiuojama:

```
a=0.3668
t=34892.34
fgg=function(x,a,t) {1/(x*gamma(a))*(x/t)^a*exp(-x/t)}
pgama=function(x){integrate(fgg,0,x)$value}
adtestas(x,pgama)
```

Gauta reikšmė: 7.973138.

Weibull skirstinio įverčiams Andersono - Darlingo testo statistika apskaičiuojama:

```
a=0.5345
t=4437.318
pweib=function(x){1-exp(-(x/t)^a)}
adtestas(x,pweib)
```

Gauta reikšmė: 4.502482.

Kadangi Pareto skirstinio pasiskirstymo funkcija taške x = 306 lygi 0, o gautas įvertis $\hat{\theta}$ = 306, tai Andersono – Darlingo testo taikyti šiam skirstiniui negalime, nes skaičiuojamas logaritmas yra neaprėžtas.

2.3.4. Chi kvadrato suderinamumo kriterijaus p - reikšmė

Bendru atveju chi kvadrato suderinamumo testo statistiką galime suskaičiuoti R paketo pagalba:

```
chitestas=function(x,func,aa,bb,cc,dd,ee,ff){
    xsrt=sort(x)
    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=func
    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 `x` - empirinė imtis, `func` - teorinė pasiskirstymo funkcija, `aa`, `bb`, `cc`, `dd`, `ee`, `ff` - intervalų galų reikšmės.

Suskaidome imtį į 7 intervalus: (0;650], (650;850], (850;1200], (1200;1700], (1700;2500], (2500;1700], (7000, ∞)

Chi kvadrato suderinamumo testo p – reikšmes randame ir statistinių lentelių.

Naudojant nupjauto lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo įverčius chi kvadrato suderinamumo testo reikšmė apskaičiuojama:

```
a=0.7635727
s=0.5407144
t=1203.249
pnlpm=function(x) {
    ifelse(x>0&x<=t, a*s*sqrt(2*pi)*pnorm(log(x/t)/s+a*s)/(exp(-(a^2)
    *(s^2)/2)+ sqrt(2*pi)*a*s*pnorm(a*s)),1-(exp(-(a^2)*(s^2)/2)
    *(t/x)^a)/(exp(-(a^2)*(s^2)/2)+ sqrt(2*pi)*a*s*pnorm(a*s)))
}
chitestas(x,pnlpm,650,850,1200,1700,2500,7000)
```

Gauta reikšmė: 0.9611185.

p – reikšmė su 3 laisvės laipsniais: 0.8107.

Naudojant sudėtinio lognormaliojo - Pareto skirstinio didžiausio tikėtinumo įverčius chi kvadrato suderinamumo testo reikšmė apskaičiuojama:

```
pslp=function(x) {
k=0.3722389
ifelse(x>0&x<=t,1/(1+pnorm(k))*pnorm((a/k)*log(x/t)+k),1-
(1/(1+pnorm(k)))*(t/x)^a)
}
a=0.7378
t=1104.371
chitestas(x,pslp,650,850,1200,1700,2500,7000)
```

Gauta reikšmė: 1.081083.

p – reikšmė su 4 laisvės laipsniais: 0.8973.

Naudojant Pareto skirstinio didžiausio tikėtinumo įverčius chi kvadrato suderinamumo testo reikšmė apskaičiuojama:

```
ppar=function(x){1-(t/x)^a}
a=0.5272
t=306
chitestas(x,ppar,650,850,1200,1700,2500,7000)
```

Gauta reikšmė: 14.41834.

p – reikšmė su 4 laisvės laipsniais: 0.0061.

Naudojant lognormaliojo skirstinio didžiausio tikėtinumo įverčius chi kvadrato suderinamumo testo reikšmė apskaičiuojama:

```
pln=function(x){pnorm((log(x)-m)/s)}
m=7.620415
s=1.3709
chitestas(x,pln,650,850,1200,1700,2500,7000)
```

Gauta reikšmė: 11,86455.

p – reikšmė su 4 laisvės laipsniais: 0,0184.

Naudojant Gama skirstinio didžiausio tikėtinumo įverčius chi kvadrato suderinamumo testo reikšmė apskaičiuojama:

```
fg=function(x) {1/(x*gamma(a))*(x/t)^a*exp(-x/t)}
pgama=function(x){integrate(fg,0,x)$value}
a=0.3668
t=34892.34
chitestas(x,pgama,650,850,1200,1700,2500,7000)
```

Gauta reikšmė: 73,81106.

p – reikšmė su 4 laisvės laipsniais: <0,001.

Naudojant Weibull skirstinio didžiausio tikėtinumo įverčius chi kvadrato suderinamumo testo reikšmė apskaičiuojama:

```
pweib=function(x){1-exp(-(x/t)^a)}
a=0.5345
t=4437.318
chitestas(x,pweib,650,850,1200,1700,2500,7000)
```

Gauta reikšmė: 37,93050.

p – reikšmė su 4 laisvės laipsniais: <0,001.

2.3.5. Sudėtiniu lognormaliuoju - Pareto skirstiniu generuotų duomenų analizės apibendrinimas

3 lentelėje yra pateikiami nupjauto lognormaliojo ir Pareto skirstinių mišinio, sudėtinio lognormaliojo – Pareto, Pareto, lognormaliojo, Gama bei Weibull skirstinių didžiausio tikėtinumo įverčiai ir su šiais įverčiais atliktų Kolmogorovo – Smirnovo, Andersono – Darlingo testų statistikų reikšmės bei chi kvadrato suderinamumo testo p – reikšmė. Kolmogorovo – Smirnovo ir Andersono – Darlingo kriterijų reikšmės yra mažiausios nupjauto lognormaliojo ir

Pareto skirstinių mišinio įverčiams. Tačiau jos neženkliai mažesnės už sudėtinio lognormaliojo – Pareto skirstinio didžiausio tikėtinumo įverčių testų rezultatų reikšmes. Mišinio įverčiams taikytų testų rezultatai yra geresni nei sudėtinio skirstinio, nes mišinys turi vienu parametru daugiau ir todėl yra lankstesnis. Chi kvadrato suderinamumo testo statistikos p – reikšmė šiek tiek didesnė sudėtinio lognormaliojo – Pareto skirstinio įverčiams nei mišinio įverčiams. Tačiau skirtumas tarp p – reikšmių nėra didelis. Kadangi chi kvadrato suderinamumo kriterijaus reikšmė priklauso nuo intervalų parinkimo, tai ir galėjo padaryti įtakos. Likusių skirstinių įverčiams taikytų kriterijų reikšmės ženkliai skiriasi nuo pirmųjų dviejų skirstinių. Kolmogorovo – Smirnovo bei Andersono – Darlingo testų statistikos yra kur kas didesnės, p – reikšmės ženkliai mažesnės. Todėl galima daryti išvadą, kad šiame skyrelyje nagrinėtą imtį geriausiai aprašo nupjauto lognormaliojo ir Pareto skirstinių mišinys bei sudėtinis lognormalusis – Pareto skirstinys.

3 lentelė

Skirstinys	**Parametrų įverčiai**	**Kolmogorovo - Smirnovo kriterijus**	**Andersono - Darlingo kriterijus**	**Chi kvadrato suderinamumo kriterijaus p - reikšmė**
Nupjauto lognormaliojo ir Pareto skirstinių mišinys	$\hat{\alpha}$ =0.7635727, $\hat{\sigma}$ =0.5407144, $\hat{\theta}$=1203.249	0.0627	0.1379	0.81
Sudėtinis lognormalusis - Pareto	$\hat{\alpha}$ =0.7377677, $\hat{\theta}$=1104.371	0.0672	0.1575	0.90
Pareto	$\hat{\alpha}$ =0.5271954, $\hat{\theta}$=306	0.2084	negalima reikšmė	0.01
Lognormalusis	$\hat{\mu}$ =7.6195839, $\hat{\sigma}$ =1.3709696	0.1346	1.7934	0.02
Gama	$\hat{\alpha}$ =0.3637933, $\hat{\theta}$=35185.4	0.3138	7.9731	<0.001
Weibull	$\hat{\tau}$ =4433.474, $\hat{\theta}$=0.5344084	0.2299	4.5024	<0.001

2.4. Draudimo nuo nelaimingų atsitikimų žalų analizė

Šiame skyrelyje nagrinėsime vienos Lietuvos ne gyvybės draudimo bendrovės draudimo nuo nelaimingų atsitikimų žalas. Draudimas nuo nelaimingų atsitikimų išmokas išmoka mirties, neįgalumo, traumos atvejais, įvykus nelaimingam atsitikimui. Taip pat mokami ligonpinigiai, kai apdraustasis guli ligoninėje ir gydomas stacionariai bei dienpinigiai, jei apdraustasis laikinai netenka darbingumo. Draudime nuo nelaimingų atsitikimų dar gali būti pasirinktos ir papildomos apsaugos nuo lengvų traumų, užsikrėtimo infekcinėmis ligomis ir pan. Nagrinėsime 809 žalas, užregistruotas per 2007 metus. Žalos svyruoja nuo 8.54 iki 200 000 Lietuvos litų.

2.4.1. Didžiausio tikėtinumo įverčiai

Naudodamiesi 2.3.1. skyrelyje aprašytomis R funkcijomis, esant pradinėms reikšmėms α =1, σ =1, θ=1000, randame nupjauto lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo įverčius:

$\hat{\alpha}$ =1.257004,

$\hat{\sigma}$ =0.831868,

$\hat{\theta}$=1263.934.

Analogiškai, su pradinėmis reikšmėmis α =1, θ=400, apskaičiuojame ir sudėtiniu lognormaliuoju – Pareto skirstinio didžiausio tikėtinumo įverčius.:

$\hat{\alpha}$ =0.727092,

$\hat{\theta}$=397.5667.

Pareto skirstinio įverčiai randami 2.3.1. skyrelyje aprašytos funkcijos pagalba:

$\hat{\alpha}$ =0.2330350,

$\hat{\theta}$=8.54.

Likusiųjų skirstinių didžiausio tikėtinumo įverčiai apskaičiuojami R paketo `MASS` funkcija `fitdistr`.

Lognormaliojo skirstinio įverčiai:

```
fitdistr(x, "lognormal", list (meanlog=0, sdlog=1))
```

Gautos reikšmės:

$\hat{\mu}$ =6.43596253,

$\hat{\sigma}$ =1.04099936.

Gama skirstinio įverčiai:

```
fitdistr(x,fg,list(a=1,t=1000),lower=0.01)
```

Čia `fg` yra Gama skirstinio tankio funkcija.

Gautos reikšmės:

$\hat{\alpha}$ =0.6207303,

$\hat{\theta}$=2706.134.

Weibull skirstinio įverčiai:

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

Gautos reikšmės:

$\hat{\tau}$ =1024.981,

$\hat{\theta}$ =0.7005034.

2.4.2. Kolmogorovo – Smirnovo kriterijus

Naudodamiesi 2.3.2 skyrelyje aprašytomis R funkcijomis, suskaičiuojame Kolmogorovo – Smirnovo testų statistikas.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant nupjauto lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo įverčius:

```
a=1.257004
s=0.831868
t=1263.934
kstestas(x, pnlpm)
```

Gauta reikšmė: 0.06534699.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant sudėtinio lognormaliojo – Pareto skirstinio didžiausio tikėtinumo įverčius:

```
a=0.727092
t=397.5667
kstestas(x, pslp)
```

Gauta reikšmė: 0.1048302.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant Pareto skirstinio didžiausio tikėtinumo įverčius:

```
a=0.2330350
t=8.54
kstestass(x,ppar)
```

Gauta reikšmė: 0.4685254.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant lognormaliojo skirstinio didžiausio tikėtinumo įverčius:

```
ks.test(x,"plnorm",6.43596253, 1.04099936)
```

Gauta reikšmė: 0.0928.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant Gama skirstinio didžiausio tikėtinumo įverčius:

```
ks.test(x,"pgamma",0.6207303,1/2706.134)
```

Gauta reikšmė: 0.2422.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant Weibull skirstinio didžiausio tikėtinumo įverčius:

```
ks.test(x,"pweibull",0.7005034, 1024.981)
```

Gauta reikšmė: 0.2207.

2.4.3. Andersono – Darlingo kriterijus

Andersono – Darlingo testo statistika atitinkamiems skirstiniams apskaičiuojama kaip ir 2.3.3. skyrelyje.

Andersono – Darlingo testo statistika nupjauto lognormaliojo ir Pareto skirstinių mišiniui:

```
a=1.257004
s=0.831868
t=1263.934
adtestas(x, pnlpm)
```

Gauta reikšmė: 3.674941.

Andersono – Darlingo testo statistika sudėtiniam lognormaliajam - Pareto skirstiniui:

```
a=0.727092
t=397.5667
adtestas(x, pslp)
```

Gauta reikšmė: 12.67835.

Andersono – Darlingo testo statistika lognormaliajam skirstiniui:

```
m=6.43596253
s=1.04099936
adtestas(x,pln)
```

Gauta reikšmė: 7.268845.

Andersono – Darlingo testo statistika Gama skirstiniui:

```
a=0.6207303
t=2706.134
adtestas(x,pgama)
```

Gauta reikšmė: 35.4672.

Andersono – Darlingo testo statistika Weibull skirstiniui:

```
a=1024.981
t=0.7005034
adtestass(x,pweib)
```

Gauta reikšmė: begalybė.

Pareto skirstinio įverčiams šio testo taikyti negalima

2.4.4. Chi kvadrato suderinamumo kriterijaus p - reikšmė

Suskaidykime sutvarkytą imtį į 7 intervalus: (0;200], (200;300], (300;400], (400;500], (500;600], (600;2000], (2000, ∞). Naudojantis 2.3.4. skyrelyje aprašytomis programomis, rasime chi kvadrato suderinamumo testų statistikas ir iš lentelių sužinosime p – reikšmes.

Nupjauto lognormaliojo ir Pareto skirstinių mišinio chi kvadrato statistika:

```
a=1.257004
s=0.831868
t=1263.934
chitestas(x, pnlpm, 200,300,400,500,600,2000)
```

Gauta reikšmė: 7.384786.

p – reikšmė su 3 laisvės laipsniais: 0.0606.

Sudėtinio lognormaliojo - Pareto skirstinio chi kvadrato statistika:

```
a=0.727092
t=397.5667
chitestas(x,pslp, 200,300,400,500,600,2000)
```

Gauta reikšmė: begalybė.

p – reikšmė su 4 laisvės laipsniais: <0.001.

Pareto skirstinio chi kvadrato statistika:

```
a=0.2330350
t=8.54
chitestas(x,ppar, 200,300,400,500,600,2000)
```

Gauta reikšmė: 1841.365.

p – reikšmė su 4 laisvės laipsniais: <0.001.

Lognormaliojo skirstinio chi kvadrato statistika:

```
m=6.43596253
s=1.04099936
chitestas(x,pln, 200,300,400,500,600,2000)
```

Gauta reikšmė: 1.251208e+17.

p – reikšmė su 4 laisvės laipsniais: <0.001.

Gama skirstinio chi kvadrato statistika:

```
a=0.6207303
t=2706.134
chitestas(x,pgama, 200,300,400,500,600,2000)
```

Gauta reikšmė: 407.933.

p – reikšmė su 4 laisvės laipsniais: <0,001.

Weibull skirstinio chi kvadrato statistika:

```
a=1024.981
t=0.7005034
chitestas(x,pweib, 200,300,400,500,600,2000)
```

Gauta reikšmė: 289.3564.

p – reikšmė su 4 laisvės laipsniais: <0,001.

2.4.5. Draudimo nuo nelaimingų atsitikimų žalų analizės apibendrinimas

4 lentelė

Skirstinys	**Parametrų įverčiai**	**Kolmogorovo - Smirnovo testo statistika**	**Andersono - Darlingo testo statistika**	**Chi kvadrato suderinamumo kriterijaus p - reikšmė**
Nupjauto lognormaliojo ir Pareto skirstinių mišinys	$\hat{\alpha}$ =1.257004, $\hat{\sigma}$ =0.831868, $\hat{\theta}$=1263.934	0.0653	3.6749	0.06
Sudėtinis lognormalusis - Pareto	$\hat{\alpha}$ =0.727092, $\hat{\theta}$=397.5667	0.1048	12.6784	<0.001
Pareto	$\hat{\alpha}$ =0.2330350, $\hat{\theta}$=8.54	0.4685	negalima reikšmė	<0.001
Lognormalusis	$\hat{\mu}$ =6.43596253, $\hat{\sigma}$ =1.04099936	0.0928	7.2688	<0.001
Gama	$\hat{\alpha}$ =0.6207303, $\hat{\theta}$=2706.134	0.2422	35.4672	<0.001
Weibull	$\hat{\tau}$ =1024.981, $\hat{\theta}$=0.7005034	0.2207	begalybė	<0.001

Lyginant rezultatus 4 lentelėje, matyti, kad mažiausios Kolmogorovo – Smirnovo ir Andersono – Darlingo testų statistikų reikšmės yra naudojant nupjautojo lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo įverčius. Visais kitais atvejais šių statistikų reikšmės yra ženkliai didesnės. Mišinio svorio konstanta *a* šiai imčiai yra 0.7941769, t.y. beveik 80% mišinio sudaro lognormalusis skirstinys. Todėl lognormaliojo skirstinio testų statistikų reikšmės yra mažesnės nei visų kitų skirstinių, išskyrus mišinį. Didžiąją dalį draudimo nuo nelaimingų atsitikimų žalų sudaro mažos žalos su dideliais dažniais, kurias gerai aprašo lognormalusis skirstinys. Chi kvadrato suderinamumo kriterijaus p – reikšmė tik mišinio atveju yra didesnė už 0.001. Nors šio testo rezultatai priklauso nuo intervalų parinkimo, tačiau, akivaizdu, jog nei vienas kitas skirstinys taip gerai neaprašo nagrinėjamu žalų, kaip nupjauto lognormaliojo ir Pareto skirstinių mišinys.

2.5. Danijos gaisrų draudimo žalų analizė

Šiame skyriuje analizuosime Danijos gaisrų draudimo žalas, įvykusias nuo 1980 iki 1990 metų. Draudimo žalų skaičius yra 2 492. Jas sudaro gaisrų padaryti nuostoliai baldams, įrenginiams, pastatams ir kitai privačiai nuosavybei. Žalos svyruoja nuo 0,3134041 iki 263,2503660 milijonų Danijos kronų.

2.5.1. Didžiausio tikėtinumo įverčiai

Naudodamiesi 2.3.1. skyrelyje aprašytomis R funkcijomis, esant pradinėms reikšmėms α =1, σ =1, θ =1, randame nupjauto lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo įverčius:

$\hat{\alpha}$ =1.328223,

$\hat{\sigma}$ =0.196517,

$\hat{\theta}$ =1.207430.

Analogiškai, su pradinėmis reikšmėmis α =1, θ =1, apskaičiuojame ir sudėtiniu lognormaliuoju – Pareto skirstinio didžiausio tikėtinumo įverčius.:

$\hat{\alpha}$ =1.436332,

$\hat{\theta}$ =1.385128.

Pareto skirstinio įverčiai randami 2.3.1. skyrelyje aprašytos funkcijos pagalba:

$\hat{\alpha}$ =0.5458171,

$\hat{\theta}$ =0.3134040.

Likusiųjų skirstinių didžiausio tikėtinumo įverčiai apskaičiuojami R paketo `MASS` funkcija `fitdistr`.

Lognormaliojo skirstinio įverčiai:

```
fitdistr(x, "lognormal", list (meanlog=0, sdlog=1))
```

Gautos reikšmės:

$\hat{\mu}$ =0.67166552,

$\hat{\sigma}$ =0.73217771

Gama skirstinio įverčiai:

```
fitdistr(x,fg,list(a=1,t=1000),lower=0.01)
```

Čia `fg` yra Gama skirstinio tankio funkcija.

Gautos reikšmės:

$\hat{\alpha}$ =1.25799429,

$\hat{\theta}$=2.43459177.

Weibull skirstinio įverčiai:

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

Gautos reikšmės:

$\hat{\tau}$ =0.94744403,

$\hat{\theta}$=2.95148311.

2.5.2. Kolmogorovo – Smirnovo kriterijus

Naudodamiesi 2.3.2 skyrelyje aprašytomis R funkcijomis, suskaičiuojame Kolmogorovo – Smirnovo testų statistikas.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant nupjauto lognormaliojo ir Pareto skirstinių mišinio didžiausio tikėtinumo įverčius:

```
a=1.328223
s=0.196517
t=1.207430
kstestas(x, pnlpm)
```

Gauta reikšmė: 0.0323.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant sudėtinio lognormaliojo – Pareto skirstinio didžiausio tikėtinumo įverčius:

```
a=1.436332
t=1.385128
kstestas(x, pslp)
```

Gauta reikšmė: 0.0287.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant Pareto skirstinio didžiausio tikėtinumo įverčius:

```
a=0.5458171
t=0.3134040
kstestass(x,ppar)
```

Gauta reikšmė: 0.4083.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant lognormaliojo skirstinio didžiausio tikėtinumo įverčius:

```
ks.test(x,"plnorm",0.67166552, 0.73217771)
```

Gauta reikšmė: 0.127.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant Gama skirstinio didžiausio tikėtinumo įverčius:

```
ks.test(x,"pgamma",1.25799429,1/2.43459177)
```

Gauta reikšmė: 0.2013.

Kolmogorovo – Smirnovo kriterijaus reikšmė naudojant Weibull skirstinio didžiausio tikėtinumo įverčius:

```
ks.test(x,"pweibull",0.94744403, 2.95148311)
```

Gauta reikšmė: 0.2556.

2.5.3. Andersono – Darlingo kriterijus

Andersono – Darlingo testo statistika atitinkamiems skirstiniams apskaičiuojama kaip ir 2.3.3. skyrelyje.

Andersono – Darlingo testo statistika nupjauto lognormaliojo ir Pareto skirstinių mišiniui:

```
a=1.328223
s=0.196517
t=1.207430
adtestas(x, pnlpm)
```

Gauta reikšmė: 3.158407.

Andersono – Darlingo testo statistika sudėtiniam lognormaliajam - Pareto skirstiniui:

```
a=1.436332
t=1.385128
adtestas(x, pslp)
```

Gauta reikšmė: 4,258509.

Andersono – Darlingo testo statistika lognormaliajam skirstiniui:

```
m=0.67166552
s=0.73217771
adtestas(x,pln)
```

Gauta reikšmė: 85,41577.

Andersono – Darlingo testo statistika Gama skirstiniui:

```
a=1.25799429
t=2.43459177
```

```
adtestas(x,pgama)
```

Gauta reikšmė: 212,447.

Andersono – Darlingo testo statistika Weibull skirstiniui:

```
a=0.94744403
t=2.95148311
adtestass(x,pweib)
```

Gauta reikšmė: begalybė.

Pareto skirstinio įverčiams šio testo taikyti negalima.

2.5.4. Chi kvadrato suderinamumo kriterijaus p - reikšmė

Suskaidykime sutvarkytą imtį į 7 intervalus: (0;0.9], (0.9;1.1], (1.1;2.3], (2.3;3.6], (3.6;4], (4;12.2], (12.2, ∞). Naudojantis 2.3.4. skyrelyje aprašytomis programomis, rasime chi kvadrato suderinamumo testų statistikas ir iš lentelių sužinosime p - reikšmes

Nupjauto lognormaliojo ir Pareto skirstinių mišinio chi kvadrato statistika:

```
a=1.328223
s=0.196517
t=1.207430
chitestas(x, pnlpm, 0.9, 1.1, 2.3, 3.6, 4, 12.2)
```

Gauta reikšmė: 1.467492.

p – reikšmė su 3 laisvės laipsniais: 0.6898.

Sudėtinio lognormaliojo - Pareto skirstinio chi kvadrato statistika:

```
a=1.436332
t=1.385128
chitestas(x,pslp, 0.9, 1.1, 2.3, 3.6, 4, 12.2)
```

Gauta reikšmė: 32.43340.

p – reikšmė su 4 laisvės laipsniais: <0.001.

Pareto skirstinio chi kvadrato statistika:

```
a=0.5458171
t=0.3134040
chitestas(x,ppar, 0.9, 1.1, 2.3, 3.6, 4, 12.2)
```

Gauta reikšmė: 3128.972.

p – reikšmė su 4 laisvės laipsniais: <0.001.

Lognormaliojo skirstinio chi kvadrato statistika:

```
m=0.67166552
s=0.73217771
chitestas(x,pln, 0.9, 1.1, 2.3, 3.6, 4, 12.2)
```

Gauta reikšmė: 878.5278.

p – reikšmė su 4 laisvės laipsniais: <0.001.

Gama skirstinio chi kvadrato statistika:

```
a=1.25799429
t=2.43459177
chitestas(x,pgama, 0.9, 1.1, 2.3, 3.6, 4, 12.2)
```

Gauta reikšmė: 1724.851.

p – reikšmė su 4 laisvės laipsniais: <0,001.

Weibull skirstinio chi kvadrato statistika:

```
a=0.94744403
t=2.95148311
chitestas(x,pweib, 0.9, 1.1, 2.3, 3.6, 4, 12.2)
```

Gauta reikšmė: 1847.144.

p – reikšmė su 4 laisvės laipsniais: <0,001.

2.5.5. Danijos žalų analizės apibendrinimas

Lyginant pasirinktų trijų kriterijų reikšmes 5 lentelėje, Kolmogorovo – Smirnovo testo rezultatai geriausi yra sudėtinio lognoramaliojo – Pareto skirstinio didžiausio tikėtinumo įverčiams, tačiau Andersono – Darlingo kriterijaus rezultatai šiek tiek geresni yra nupjauto lognormaliojo ir Pareto skirstinių mišinio parametrų įverčiams. Abiejų testų rezultatai lyginant mišinio ir sudėtinio skirstinio parametrus skiriasi neženkliai. Chi kvadrato suderinamumo testo p – reikšmė akivaizdžiai didžiausia yra mišinio atveju. Visais kitais atvejais yra mažesnė už 0.001. Taigi, galime daryti išvadą, kad geriausiai Danijos gaisrų draudimo žalų duomenis aprašo skirstinių mišinys. Mišinio svorio konstanta šiuo atveju yra lygi 0.2898601, t.y. nepilnai 30% mišinio svorio sudaro lognormalusis skirstinys. Tai lemia labai didelės žalų išskirstys. Kitų skirstinių didžiausio tikėtinumo įverčiams pritaikytų kriterijų reikšmės yra keliasdešimt kartų didesnės, todėl naudoti juos, vertinant Danijos gaisrų žalas, yra netikslinga.

5 lentelė

Skirstinys	**Parametrų įverčiai**	**Kolmogorovo - Smirnovo testo statistika**	**Andersono - Darlingo testo statistika**	**Chi kvadrato suderinamumo kriterijaus p - reikšmė**
Nupjauto lognormaliojo ir Pareto skirstinių mišinys	$\hat{\alpha}$ =1.328223, $\hat{\sigma}$ =0.196517, $\hat{\theta}$=1.207430	0.0323	3.1584	0.6898
Sudėtinis lognormalusis - Pareto	$\hat{\alpha}$ =1.436332, $\hat{\theta}$=1.385128	0.0287	4.2585	<0.001
Pareto	$\hat{\alpha}$ =0.5458171, $\hat{\theta}$=0.3134040	0.4083	negalima reikšmė	<0.001
Lognormalusis	$\hat{\mu}$ =0.67166552, $\hat{\sigma}$ =0.73217771	0.127	85.4157	<0.001
Gama	$\hat{\alpha}$ =1.25799429, $\hat{\theta}$=2.43459177	0.2013	212.447	<0.001
Weibull	$\hat{\tau}$ =0.94744403, $\hat{\theta}$=2.95148311	0.2556	begalybė	<0.001

Išvados

Darbe parodėme, kaip sudarius dviejų skirstinių mišinį, gauname modelį, tinkanti įvairių tipų draudimo žaloms vertinti. Nors lognormalusis skirstinys su dviem parametrais tinka modeliuoti mažas žalas su dideliais dažniais, o Pareto skirstinys su dviem parametrais, tinkamesnis didelėms žaloms su mažais dažniais, mišinys su trimis parametrais gerai vertina visą draudimo žalų imtį nuo mažųjų iki labai didelių. Nupjauto lognormaliojo ir Pareto skirstinių mišinio tankio funkcijos forma yra panaši į lognormaliojo skirstinio, tačiau jo viršutinė uodega yra didesnė. Mišinys yra tinkamesnis, kai lognormalusis skirstinys nepakankamai įvertina dideles žalas. Sudėtinis lognormalusis – Pareto skirstinys taip pat tinkamas visų dydžių žaloms analizuoti, tačiau jis nepakankamai gerai įvertina draudimo žalas, kai lognormaliojo skirstinio dalis mišinyje yra daugiau nei 40 - 50%, t.y. nepakankamai tiksliai įvertina imtį tuomet, kai didžioji dalis žalų yra mažos su dideliais dažniais.

Išanalizavus trijų rūšių duomenis – generuotus sudėtiniu lognormaliuoju – Pareto skirstiniu, draudimo nuo nelaimingų atsitikimų ir Danijos gaisrų draudimo žalas, įsitikinome, kad nupjauto lognormaliojo ir Pareto skirstinių mišinys puikiai tinka realioms draudimo žaloms vertinti. Kai lognormaliojo skirstinio mišinyje yra mažiau nei 40 - 50%, tuomet žalas pakankamai gerai vertina ir sudėtinis lognormalusis – Pareto skirstinys, tačiau, esant mažesnei daliai Pareto skirstinio mišinyje, sudėtinio skirstinio įverčiai yra prastesni nei mišinio. Tam įtakos turi paties sudėtinio lognormaliojo – Pareto skirstinio struktūra, nes normalizuojanti konstanta jame yra fiksuota ir negali laisvai kisti, kitaip nei nupjauto lognormaliojo ir Pareto skirstinių mišinio atveju. Mišinio svorio konstanta priklauso nuo dviejų parametrų α ir σ, todėl kinta, priklausomai nuo parametrų įverčių ir duoda tikslesnius rezultatus.

Taigi, nupjauto lognormaliojo ir Pareto skirstinių mišinys gerai vertina tiek imtis, kuriose yra daug mažų žalų su dideliais dažniais, tiek tas, kuriose yra daug didelių išskirčių.

Literatūros sąrašas

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

2. V. Čekanavičius, G. Murauskas. *Statistika ir jis taikymai. I*, Vilnius: TEV, 2000.

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

4. G. Kuodis, G. Tamulytė, V. Žuklijaitė. *Žalų modeliavimas naudojant sudėtinį lognormalųjį – Pareto skirstinį*, Bakalauro darbas, 2006.

5. S. I. Resnick. Discussion of the Danish data on large fire insurance loss, *ASTIN Bulletin*, 1997, 27, p. 139-151.