

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MATEMAATIKA JA STATISTIKA INSTITUUT

Hardo Niit

# **Kaskokindlustuse kahjude modelleerimine**

Kindlustus- ja finantsmatemaatika

Magistritöö (30 EAP)

Juhendaja: emeriitprofessor Tõnu Kollo

TARTU 2021

# KASKOKINDLUSTUSE KAHJUDE MODELLEERIMINE

Magistritöö

Hardo Niit

## Lühikokkuvõte

Käesoleva magistritöö eesmärk on kaskokindlustuse kahjuandmete modelleerimine kasutades erinevaid tõenäosusjaotusi. Jaotuste sobivuse hindamiseks kasutatakse mitmeid meetodeid nagu Kolmogorov-Smirnovi test,  $\chi^2$ -test, jääkkeskmise meetod ning kvantiil-kvantiil graafikuid. Töö esimeses osas antakse ülevaade kasutatud kahjujaotustest. Töö teises osas antakse ülevaade jääkkeskmise meetodist ning tuletatakse jääkkeskmise valemid log-normaalse ja nihutatud Pareto jaotuse korral. Kolmandas osas antakse ülevaade andmes- tikust, Kolmogorov-Smirnovi ja  $\chi^2$ -testidest ning sobitatakse jaotusi kahju- andmetele.

**CERCS teaduseriala:** P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

**Märksõnad:** Kahjujaotused, Kolmogorov-Smirnovi teststatistik, hii-ruut test- statistik, jääkkeskmise meetod.

# MODELLING CASCO INSURANCE CLAIMS

Master thesis

Hardo Niit

## Abstract

The aim of this Master thesis is to fit various loss distributions to the casco insurance claims and assess their fit to data. For goodness-of-fit testing various methods are used including Kolmogorov-Smirnov test,  $\chi^2$ -test, mean excess method and quantile-quantile plots. In the first part, some well-known loss distributions and their characteristics are presented. In the second part overview of the mean excess method is given and formulas of the mean excess for the log-normal and shifted Pareto distribution are derived. In the third part an overview of the data is given, used methods for goodness-of-fit testing are described and actual fitting of the distributions to data is carried out.

**CERCS research specialisation:** P160 Statistics, operations research, programming, financial and actuarial mathematics.

**Key Words:** Loss distributions, Kolmogorov-Smirnov test, chi-square test, excess of loss.

# Sisukord

<b>Sissejuhatus</b>	<b>5</b>
<b>1 Jaotused</b>	<b>7</b>
1.1 Gammajaotus . . . . .	7
1.2 Gaussi pöördjaotus (Waldi jaotus) . . . . .	8
1.3 Log-normaalne jaotus . . . . .	10
1.4 Nihutatud Pareto jaotus . . . . .	11
1.5 Weibulli jaotus . . . . .	19
<b>2 Jääkkeskmise meetod (<i>Mean excess method</i>)</b>	<b>23</b>
2.1 Meetodi tõenäosuslik kirjeldus . . . . .	23
2.2 Pareto ja log-normaalne jaotuse jääkkeskmised . . . . .	24
<b>3 Kahjude analüüs</b>	<b>30</b>
3.1 Andmete kirjeldus . . . . .	30
3.2 Jaotuste sobitamine kahjuandmetele . . . . .	33
3.3 Jaotuste sobitamine kogu kahjuandmestikule . . . . .	35
3.4 Andmestiku jagamine kaheks kahjusuuruse kaudu . . . . .	40
3.4.1 Jaotuste sobitamine kahjuandmestiku esiosale . . . . .	41
3.4.2 Jaotuste sobitamine kahjuandmestiku sabaosale . . . . .	44
3.5 Andmestiku jagamine kaheks kahjuliigiti . . . . .	48
3.5.1 Jaotuste sobitamine esiklaasikahjudele . . . . .	49
3.5.2 Jaotuste sobitamine ülejäänud kahjudele . . . . .	52
<b>4 Kokkuvõte</b>	<b>57</b>

<b>5</b>	<b>Kasutatud kirjandus</b>	<b>58</b>
<b>6</b>	<b>Lisad</b>	<b>59</b>
6.1	Kvantiilide tabelid . . . . .	59
6.2	Kolmogorov-Smirnovi teststatistikute varieeruvus . . . . .	60

## Sissejuhatus

Iga tegevusega, mis me ette võtame, kaasneb risk, et tulemus ei ole see, mida me soovime. Sõidukid on saanud meie argielu lahutamatuks osaks ning osaledes liikluses esineb alati risk, et võib juhtuda õnnetus. Üheks võimaluseks leevendada riski, on see kindlustada ja seda võimalust pakuvad mitmed erinevad kindlustusseltsid. Kaskokindlustus on üks kindlustuse liik, mis hoiab õnnetuse korral ära suure rahalise kahju kindlustusvõtja jaoks ning aitab vältida ebameeldivaid ootamusi.

Samas on vaja ka kindlustusseltsil ehk kindlustusandjal toime tulla võetud riskidega. Kuna kahjud on oma olemuselt juhuslikud suurused ja toimuvad juhuslikel ajamomentidel, siis nende suurust ja toimumise hetke on raske ette ennustada. Sellest tekib vajadus hinnata võimalikke jaotusi, mis kirjeldavad juba olemasolevaid kahjuandmeid. Olles leidnud kahjuandmeid kirjeldava jaotuse, on võimalik anda hinnanguid tulevikus tekkivate kahjude suurustele.

Magistritöö eesmärgiks on sobitada mitmeid jaotusi If P&C Insurance AS kaskokindlustuse kahjuandmetele ning hinnata nende sobivust kasutades jääkkeskmise meetodit,  $\chi^2$ -testi, Kolmogorov-Smirnovi testi ning kvantiil-kvantiil graafikuid.

Töö esimeses peatükis anname ülevaate tuntumatest jaotustest, mida kasutatakse kindlustuskahjude modelleerimiseks. Toome välja jaotuste tihedus- ja jaotusfunktsioonid ning jaotuste põhilised arvarakteristikud. Tuletame kahe jaotuse korral nende asümmeetria- ja järsakuskordaja.

Töö teises peatükis anname ülevaate jääkkeskmise meetodist. Toome ära jääkkeskmise arvutamise valemi ning esitame jääkkeskmise empiirilise väärtuse leidmise eeskirja. Tuletame jääkkeskmise valemi Pareto ning log-normaalse jaotuse jaoks. Lisaks uurime log-normaalse jaotuse jääkkeskmise kuju sõltuvust parameetritest.

Kolmandas peatükis anname ülevaate töös kasutusel olevast andmestikust, tutvustame kahte enim tuntud meetodit jaotuste sobivuse hindamiseks. Sobitame jaotusi kahjude empiirilisele jaotusele ning toome välja kvantiil-kvantiil graafikud, jääkeskmise graafikud ning Kolmogorov-Smirnovi ja  $\chi^2$ -teststatistikute väärtused. Algselt sobitame jaotusi kogu kahjuandmestikule ning seejärel jagame andmed kahel erineval viisil kaheks. Esiteks jagame kahjud suuruse järgi kaheks ning sobitame jaotusi eraldi kahjuandmete esiosal ning sabaosal. Teiseks eraldame kogu kahjuandmestikust esiklaasikahjud ning sobitame jaotusi eraldi esiklaasikahjudele ja ülejäänud kahjudele.

Töö lisas on välja toodud tabelid kvantiilide võrdlemise kohta erinevate kahjuandmete ning jaotuste korral. Lisaks on ära toodud tabelid, mis näitavad Kolmogorov-Smirnovi teststatistiku varieeruvust erinevate kahjuandmete korral.

Magistritöö kirjutamisel on kasutatud programmi  $\text{\LaTeX}$  ning töös teostatud analüüs ja graafikud on koostatud statistikatarkvaras R (versioon 3.6.3).

Siinkohal kasutab autor võimalust tänada oma juhendajat emeriitprofessor Tõnu Kollot asjakohaste nõuannete ja ideede eest.

# 1 Jaotused

Kindlustuskahjude puhul on tegemist mittenegatiivsete juhuslike suuruste väärtustega, seega kasutatakse kindlustuskahjude modelleerimiseks pidevaid jaotusi, mis on defineeritud arvtelje mittenegatiivsel osal. Antud magistritöö raames vaatleme viit jaotust ning toome välja nende peamised arvkarakteristikud. Nendeks on lognormaalne, Weibulli, Pareto, Waldi ning gammajaotus. Waldi jaotust tuntakse ka Gaussi pöördjaotusena. Jaotused on esitatud raamatute (Forbes *et al.*, 2011; Tse, 2009) põhjal. Täiendavate allikate puhul on neile tekstis täiendavalt viidatud.

## 1.1 Gammajaotus

**Definitsioon. 1.1.** *Öeldakse, et juhuslik suurus  $X$  on gammajaotusega kujuparameetriga  $\alpha$  ja skaalaparameetriga  $\lambda$ ,  $X \sim \Gamma(\alpha, \lambda)$ , kus  $\alpha > 0$  ja  $\lambda > 0$ , kui tema tihedusfunktsioon avaldub kujul*

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\lambda}, \quad x \geq 0. \quad (1)$$

Siin  $\Gamma(\alpha)$  on gammafunktsioon, mis on defineeritud järgmiselt

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0. \quad (2)$$

Juhul, kui  $\alpha > 1$ , siis  $\Gamma(\alpha)$  on esitatav rekursiooni  $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$  teel.

Kui  $\alpha$  on positiivne täisarv, siis  $\Gamma(\alpha) = (\alpha - 1)!$ .

Gammajaotuse jaotusfunktsioon avaldub järgmisel kujul

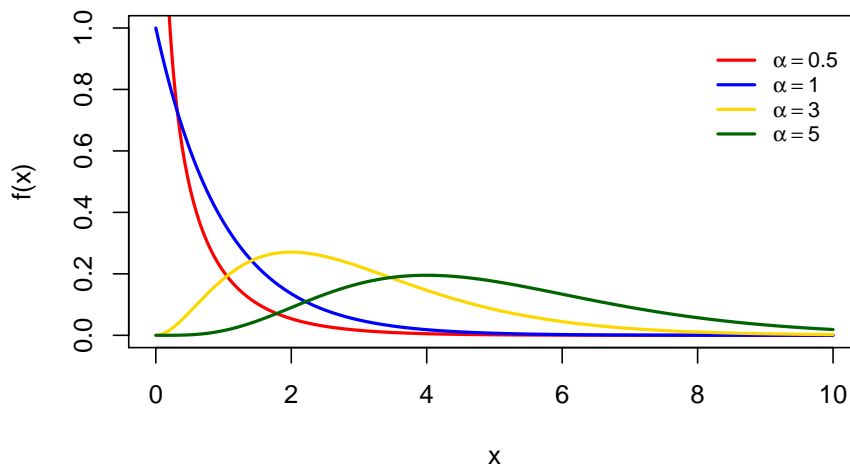
$$F_X(x) = \frac{\gamma(\alpha, x\lambda)}{\Gamma(\alpha)},$$

kus  $\gamma(\alpha, t)$  on mittetäielik gammafunktsioon,

$$\gamma(\alpha, t) = \int_0^t x^{\alpha-1} e^{-x} dx.$$

Gammajaotusega juhusliku suuruse  $X$  põhilised arvkarakteristikud on järgmised:

- keskväärtus:  $EX = \frac{\alpha}{\lambda}$ ;
- dispersioon:  $DX = \frac{\alpha}{\lambda^2}$ ;
- asümmeetriakordaja:  $\beta_1(X) = \frac{2}{\sqrt{\alpha}}$ ;
- järsakuskordaja:  $\beta_2(X) = 3 + \frac{6}{\alpha}$ .



Joonis 1: Gammajaotuse tihedusfunktsioon, kui  $\lambda = 1$ .

Juhul, kui  $\alpha = 1$ , siis  $\Gamma(1) = 1$  ja valem (1) taandub kujule  $f_X(x) = \lambda e^{-x\lambda}$ , mis on eksponentjaotuse tihedusfunktsioon parameetriga  $\lambda$  ehk eksponentjaotus on gammajaotuse erijuht.

## 1.2 Gaussi pöördjaotus (Waldi jaotus)

**Definitsioon. 1.2.** Öeldakse, et juhuslik suurus  $X$  on Waldi jaotusega kohaparametriga  $\mu$  ja skaalaparametriga  $\lambda$ ,  $X \sim IG(\mu, \lambda)$ , kus  $\mu > 0$  ja  $\lambda > 0$ , kui tema

tihedusfunktsioon avaldub kujul

$$f_X(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x \geq 0.$$

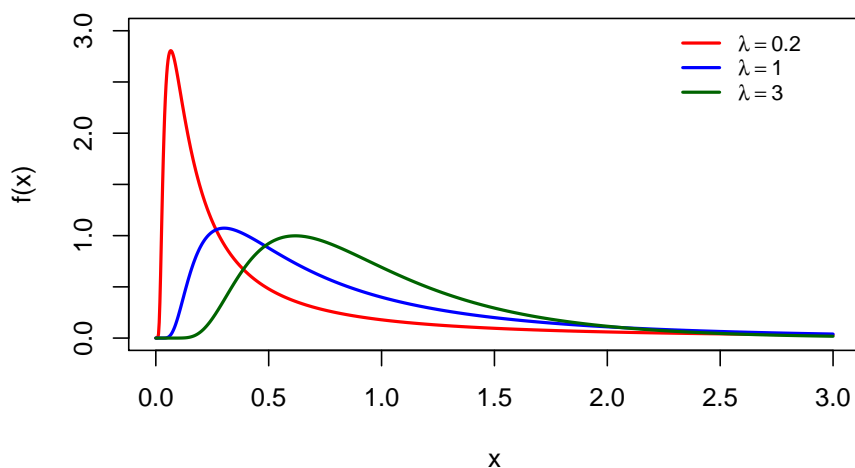
Waldi jaotuse jaotusfunktsioon avaldub järgmisel kujul

$$F_X(x) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right),$$

kus  $\Phi(\cdot)$  on standardse normaaljaotuse jaotusfunktsioon.

Waldi jaotusega juhusliku suuruse  $X$  arvkarakteristikud on järgmised:

- keskväärtus:  $EX = \mu$ ;
- dispersioon:  $DX = \frac{\mu^3}{\lambda}$ ;
- asümmeetriakordaja:  $\beta_1(X) = 3\sqrt{\frac{\mu}{\lambda}}$ ;
- järsakuskordaja:  $\beta_2(X) = 3 + 15\frac{\mu}{\lambda}$ .



Joonis 2: Waldi jaotuse tihedusfunktsioon, kui  $\mu = 1$ .

### 1.3 Log-normaalne jaotus

**Definitsioon. 1.3.** *Öeldakse, et juhuslik suurus  $X$  on log-normaalne jaotusega kohaparametri  $\mu$  ja kujuparametri  $\sigma$ ,  $X \sim LN(\mu, \sigma)$ , kus  $-\infty < \mu < \infty$  ja  $\sigma > 0$ , kui tema tihedusfunktsioon avaldub kujul*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad x \geq 0.$$

Log-normaalne jaotuse jaotusfunktsioon avaldub järgmisel kujul

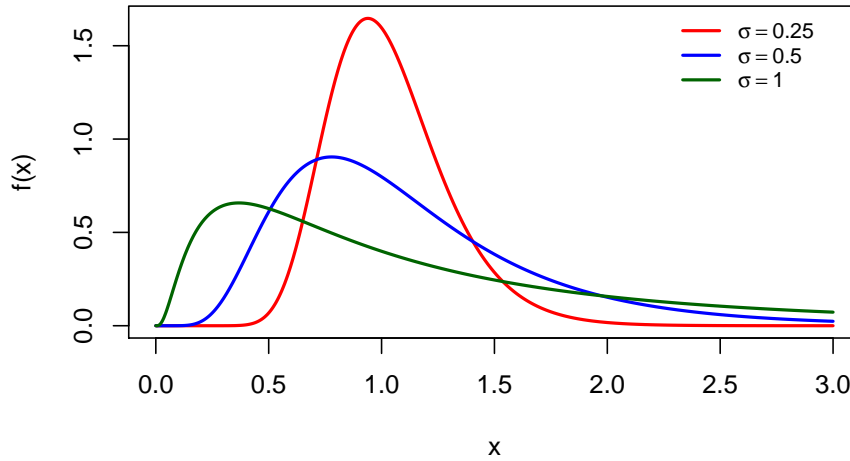
$$F_X(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right),$$

kus  $\Phi(\cdot)$  on standardse normaaljaotuse jaotusfunktsioon.

Kui juhuslik suurus  $X$  on log-normaalne jaotusega, siis juhuslik suurus  $Y = \ln X$  on normaaljaotusega ja vastupidi, kui juhuslik suurus  $Y$  on normaaljaotusega, siis  $X = e^Y$  on log-normaalne jaotusega. Log-normaalne jaotuse tõenäosustihedus on asümmeetriline ja positiivse kaldega.

Log-normaalne jaotusega juhusliku suuruse  $X$  arvkarakteristikud on järgmised:

- keskvärtus:  $EX = e^{\mu + \frac{\sigma^2}{2}}$ ;
- dispersioon:  $DX = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$ ;
- asümmeetriakordaja:  $\beta_1(X) = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$ ;
- järsakuskordaja:  $\beta_2(X) = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3$ .



Joonis 3: Log-normaalse jaotuse tihedusfunktsioon, kui  $\mu = 0$ .

## 1.4 Nihutatud Pareto jaotus

**Definitsioon. 1.4.** *Öeldakse, et juhuslik suurus  $X$  on nihutatud Pareto jaotusega (kirjanduses tuntud ka kui Ameerika Pareto jaotus või nullpunkti nihutatud Pareto jaotus) kohaparametri  $\alpha$  ja kujuparametri  $\beta$ ,  $X \sim Pa(\alpha, \beta)$ , kus  $\alpha > 0$  ja  $\beta > 0$ , kui tema tihedusfunktsioon avaldub kujul*

$$f_X(x) = \frac{\alpha\beta^\alpha}{(x + \beta)^{\alpha+1}}, \quad x \geq 0.$$

Nihutatud Pareto jaotuse jaotusfunktsioon avaldub järgmisel kujul

$$F_X(x) = 1 - \left(\frac{\beta}{x + \beta}\right)^\alpha. \quad (3)$$

Nihutatud Pareto (edaspidi nimetame Pareto) jaotusega juhusliku suuruse  $X$  arv-  
karakteristikud on järgmised:

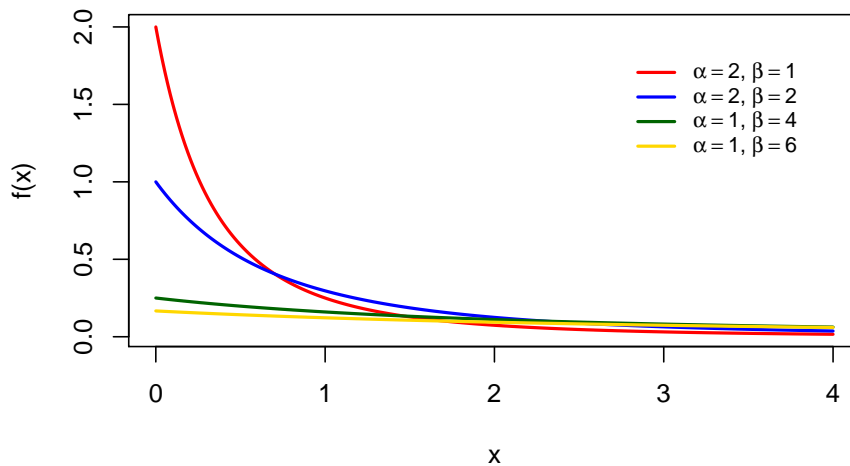
- keskväärtus:  $EX = \frac{\beta}{\alpha-1}$ ;

- dispersioon:  $DX = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}$ ; (4)

- asümmeetriakordaja:  $\beta_1(X) = \frac{2(\alpha+1)}{\alpha-3} \sqrt{\frac{\alpha-2}{\alpha}}$ ; (5)

- järsakuskordaja:  $\beta_2(X) = \frac{3(3\alpha^3 - 5\alpha^2 - 4)}{\alpha(\alpha-3)(\alpha-4)}$ . (6)

Kuna eespool viidatud raamatutes Forbes et. al (2011), Tse (2009) ei ole ära toodud selle jaotuse asümmeetriakordaja  $\beta_1(X)$  ja järsakuskordaja  $\beta_2(X)$  avaldised, siis võrdused (5) ja (6) on järgnevalt leitud pärast vajalike momentide leidmist (Laused 1.3 ja 1.4).



Joonis 4: Pareto jaotuse tihedusfunktsioon erinevate parameetrite korral.

Järgmiseks tuletame selle jaotuse eelnevalt toodud asümmeetriakordaja ja järsakuskordaja momentide kaudu.

**Lause 1.1.** Pareto jaotuse esimesed neli momenti on järgmisel kujul

$$\begin{aligned}
 EX &= \frac{\beta}{(\alpha - 1)}; \\
 EX^2 &= \frac{2\beta}{(\alpha - 1)(\alpha - 2)}; \\
 EX^3 &= \frac{6\beta}{(\alpha - 1)(\alpha - 2)(\alpha - 3)}; \\
 EX^4 &= \frac{24\beta}{(\alpha - 1)(\alpha - 2)(\alpha - 3)(\alpha - 4)}.
 \end{aligned}$$

**Tõestus:** Leiame esmalt Pareto jaotuse momendid  $E(X + \beta)^k$ , kus  $k = 1, \dots, 4$  ja avaldame sealt seejärel momendid.

Leiame esimest järku momendi  $E(X + \beta)$

$$\begin{aligned}
 E[(X + \beta)] &= \int_0^\infty (\beta + x) \frac{\alpha\beta^\alpha}{(\beta + x)^{\alpha+1}} dx = \alpha\beta^\alpha \int_0^\infty \frac{(\beta + x)}{(\beta + x)^{\alpha+1}} dx \\
 &= \alpha\beta^\alpha \int_0^\infty (\beta + x)^{-\alpha} dx = \alpha\beta^\alpha \left. \frac{(\beta + x)^{-\alpha+1}}{-\alpha + 1} \right|_0^\infty \\
 &= \alpha\beta^\alpha \left( \frac{(\beta + \infty)^{-\alpha+1}}{-\alpha + 1} - \frac{(\beta + 0)^{-\alpha+1}}{-\alpha + 1} \right) = \alpha\beta^\alpha \left( \frac{\beta^{1-\alpha}}{\alpha - 1} \right) \\
 &= \frac{\alpha\beta}{\alpha - 1}.
 \end{aligned}$$

Keskväärtuse saame kujul

$$\begin{aligned}
 EX &= E[(X + \beta) - \beta] = E[(X + \beta)] - \beta = \frac{\alpha\beta}{\alpha - 1} - \beta \\
 &= \frac{\alpha\beta - (\alpha - 1)\beta}{\alpha - 1} = \frac{\alpha\beta - \alpha\beta + \beta}{\alpha - 1} = \frac{\beta}{(\alpha - 1)}.
 \end{aligned}$$

Leiame teist järku momendi  $E(X + \beta)^2$ .

$$\begin{aligned}
 E[(X + \beta)^2] &= \int_0^\infty (\beta + x)^2 \frac{\alpha\beta^\alpha}{(\beta + x)^{\alpha+1}} dx = \alpha\beta^\alpha \int_0^\infty \frac{(\beta + x)^2}{(\beta + x)^{\alpha+1}} dx \\
 &= \alpha\beta^\alpha \int_0^\infty (\beta + x)^{-\alpha+1} dx = \alpha\beta^\alpha \frac{(\beta + x)^{-\alpha+2}}{-\alpha+2} \Big|_0^\infty \\
 &= \alpha\beta^\alpha \left( \frac{(\beta + \infty)^{-\alpha+2}}{-\alpha+2} - \frac{(\beta + 0)^{-\alpha+2}}{-\alpha+2} \right) = \alpha\beta^\alpha \left( \frac{\beta^{2-\alpha}}{\alpha-2} \right) \\
 &= \frac{\alpha\beta^2}{\alpha-2}.
 \end{aligned}$$

Siit saame teist järku momendi

$$\begin{aligned}
 EX^2 &= E[(X + \beta)^2 - 2\beta X - \beta^2] = E[(X + \beta)^2] - 2\beta EX - \beta^2 \\
 &= \frac{\alpha\beta^2}{\alpha-2} - 2\beta \frac{\beta}{\alpha-1} - \beta^2 = \frac{(\alpha-1)\alpha\beta^2 - (\alpha-2)2\beta^2 - (\alpha^2 - 3\alpha + 2)\beta^2}{(\alpha-1)(\alpha-2)} \\
 &= \frac{\alpha^2\beta^2 - \alpha\beta^2 - 2\alpha\beta^2 + 4\beta^2 - \alpha^2\beta^2 + 3\alpha\beta^2 - 2\beta^2}{(\alpha-1)(\alpha-2)} \\
 &= \frac{2\beta^2}{(\alpha-1)(\alpha-2)}.
 \end{aligned}$$

Järgnevalt leiame kolmandat järku momendi  $E(X + \beta)^3$ .

$$\begin{aligned}
 E[(X + \beta)^3] &= \int_0^\infty (\beta + x)^3 \frac{\alpha\beta^\alpha}{(\beta + x)^{\alpha+1}} dx = \alpha\beta^\alpha \int_0^\infty \frac{(\beta + x)^3}{(\beta + x)^{\alpha+1}} dx \\
 &= \alpha\beta^\alpha \int_0^\infty (\beta + x)^{-\alpha+2} dx = \alpha\beta^\alpha \frac{(\beta + x)^{-\alpha+3}}{-\alpha+3} \Big|_0^\infty \\
 &= \alpha\beta^\alpha \left( \frac{(\beta + \infty)^{-\alpha+3}}{-\alpha+3} - \frac{(\beta + 0)^{-\alpha+3}}{-\alpha+3} \right) = \alpha\beta^\alpha \left( \frac{\beta^{3-\alpha}}{\alpha-3} \right) \\
 &= \frac{\alpha\beta^3}{\alpha-3}.
 \end{aligned}$$

Kolmandat järku momendi saame pärast teisendusi

$$\begin{aligned}
 EX^3 &= E[(X + \beta)^3 - 3\beta X^2 - 3\beta^2 X - \beta^3] \\
 &= E[(X + \beta)^3] - 3\beta EX^2 - 3\beta^2 EX - \beta^3 \\
 &= \frac{\alpha\beta^3}{\alpha - 3} - 3\beta \frac{2\beta^2}{(\alpha - 1)(\alpha - 2)} - 3\beta^2 \frac{\beta}{(\alpha - 1)} - \beta^3 \\
 &= \frac{(\alpha^2 - 3\alpha + 2)\alpha\beta^3 - (\alpha - 3)6\beta^3 - (\alpha^2 - 5\alpha + 6)3\beta^3 - (\alpha^3 - 6\alpha^2 + 11\alpha - 6)\beta^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} \\
 &= \frac{\alpha^3\beta^3 - 3\alpha^2\beta^3 + 2\alpha\beta^3 - 6\alpha\beta^3 + 18\beta^3 - 3\alpha^2\beta^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} \\
 &+ \frac{15\alpha\beta^3 - 18\beta^3 - \alpha^3\beta^3 + 6\alpha^2\beta^3 - 11\alpha\beta^3 + 6\beta^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} \\
 &= \frac{6\beta^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)}.
 \end{aligned}$$

Leiame neljandat järku momendi  $E(X + \beta)^4$ .

$$\begin{aligned}
 E[(X + \beta)^4] &= \int_0^\infty (\beta + x)^4 \frac{\alpha\beta^\alpha}{(\beta + x)^{\alpha+1}} dx = \alpha\beta^\alpha \int_0^\infty \frac{(\beta + x)^4}{(\beta + x)^{\alpha+1}} dx \\
 &= \alpha\beta^\alpha \int_0^\infty (\beta + x)^{-\alpha+3} dx = \alpha\beta^\alpha \frac{(\beta + x)^{-\alpha+4}}{-\alpha + 4} \Big|_0^\infty \\
 &= \alpha\beta^\alpha \left( \frac{(\beta + \infty)^{-\alpha+4}}{-\alpha + 4} - \frac{(\beta + 0)^{-\alpha+4}}{-\alpha + 4} \right) = \alpha\beta^\alpha \left( \frac{\beta^{4-\alpha}}{\alpha - 4} \right) \\
 &= \frac{\alpha\beta^4}{\alpha - 4}.
 \end{aligned}$$

Neljandat järku momendi leiame samal põhimõttel

$$\begin{aligned}
EX^4 &= E[(X + \beta)^4 - 4\beta X^3 - 6\beta^2 X^2 - 4\beta^3 X - \beta^4] \\
&= E[(X + \beta)^4] - 4\beta EX^3 - 6\beta^2 EX^2 - 4\beta^3 EX - \beta^4 \\
&= \frac{\alpha\beta^4}{\alpha - 4} - 4\beta \frac{6\beta^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} - 6\beta^2 \frac{2\beta^2}{(\alpha - 1)(\alpha - 2)} - 4\beta^3 \frac{\beta}{(\alpha - 1)} - \beta^4 \\
&= \frac{(\alpha^3 - 6\alpha^2 + 11\alpha - 6)\alpha\beta^4 - (\alpha - 4)24\beta^4 - (\alpha^2 - 7\alpha + 12)12\beta^4}{(\alpha - 1)(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&+ \frac{(\alpha^3 - 9\alpha^2 + 26\alpha - 24) - 4\beta^4 - (\alpha^4 - 10\alpha^3 + 35\alpha^2 - 50\alpha + 24)\beta^4}{(\alpha - 1)(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&= \frac{\alpha^4\beta^4 - 6\alpha^3\beta^4 + 11\alpha^2\beta^4 - 6\alpha\beta^4 - 24\alpha\beta^4 + 96\beta^4 - 12\alpha^2\beta^4 + 84\alpha\beta^4 - 144\beta^4}{(\alpha - 1)(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&+ \frac{-4\alpha^3\beta^4 + 36\alpha^2\beta^4 - 104\alpha\beta^4 + 96\beta^4 - \alpha^4\beta^4 + 10\alpha^3\beta^4 - 35\alpha^2\beta^4 + 50\alpha\beta^4 - 24\beta^4}{(\alpha - 1)(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&= \frac{24\beta^4}{(\alpha - 1)(\alpha - 2)(\alpha - 3)(\alpha - 4)}.
\end{aligned}$$

**Lause 1.2.** Pareto jaotuse kolmas ja neljas tsentraalne moment on järgmisel kujul

$$E[(X - EX)^3] = \frac{2\beta^3\alpha(\alpha + 1)}{(\alpha - 1)^3(\alpha - 2)(\alpha - 3)}; \quad (7)$$

$$E[(X - EX)^4] = \frac{3\beta^4\alpha(3\alpha^2 + \alpha + 2)}{(\alpha - 1)^4(\alpha - 2)(\alpha - 3)(\alpha - 4)}. \quad (8)$$

**Tõestus:** Leiame nimetatud Pareto jaotuse tsentraalsed momendid. Esiteks leiame kolmandat järku tsentraalse momendi

$$\begin{aligned}
E[(X - EX)^3] &= E(X^3 - 3X^2EX + 3X(EX)^2 - (EX)^3) \\
&= E(X^3) - 3EXE(X^2) + 2(EX)^3 \\
&= \frac{6\beta^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} - 3\frac{\beta}{(\alpha - 1)}\frac{2\beta^2}{(\alpha - 1)(\alpha - 2)} + \frac{2\beta^3}{(\alpha - 1)^3} \\
&= \frac{6\beta^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} - \frac{6\beta^3}{(\alpha - 1)^2(\alpha - 2)} + \frac{2\beta^3}{(\alpha - 1)^3} \\
&= \frac{6\beta^3(\alpha - 1)^2 - 6\beta^3(\alpha - 1)(\alpha - 3) + 2\beta^3(\alpha - 2)(\alpha - 3)}{(\alpha - 1)^3(\alpha - 2)(\alpha - 3)}
\end{aligned}$$

$$\begin{aligned}
&= \frac{12\beta^3(\alpha - 1) + 2\beta^3(\alpha - 2)(\alpha - 3)}{(\alpha - 1)^3(\alpha - 2)(\alpha - 3)} \\
&= \frac{2\beta^3\alpha(\alpha + 1)}{(\alpha - 1)^3(\alpha - 2)(\alpha - 3)}.
\end{aligned}$$

Neljandat järku tsentraalse momendi tuletuskäik on järgnev

$$\begin{aligned}
E[(X - EX)^4] &= E(X^4 - 4X^3EX + 6X^2(EX)^2 - 4X(EX)^3 + (EX)^4) \\
&= E(X^4) - 4E(X^3)EX + 6E(X^2)(EX)^2 - 3(EX)^4 \\
&= \frac{24\beta^4}{(\alpha - 1)(\alpha - 2)(\alpha - 3)(\alpha - 4)} - \frac{24\beta^4}{(\alpha - 1)^2(\alpha - 2)(\alpha - 3)} \\
&\quad + \frac{12\beta^4}{(\alpha - 1)^3(\alpha - 2)(\alpha - 3)} - \frac{3\beta^4}{(\alpha - 1)^4} \\
&= \frac{24\beta^4(\alpha - 1)^3 - 24\beta^4(\alpha - 1)^2(\alpha - 4)}{(\alpha - 1)^4(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&\quad + \frac{12\beta^4(\alpha - 1)(\alpha - 3)(\alpha - 4) - 3\beta^4(\alpha - 2)(\alpha - 3)(\alpha - 4)}{(\alpha - 1)^4(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&= \frac{3\beta^4(8(\alpha - 1)^2(\alpha - 1 - \alpha + 4) + (\alpha - 3)(\alpha - 4)(4(\alpha - 1) - \alpha + 2))}{(\alpha - 1)^4(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&= \frac{3\beta^4(24(\alpha - 1)^2 + (\alpha^2 - 7\alpha + 12)(3\alpha - 2))}{(\alpha - 1)^4(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&= \frac{3\beta^4(24\alpha^2 - 48\alpha + 24 + 3\alpha^3 - 21\alpha^2 + 36\alpha - 2\alpha^2 + 14\alpha - 24)}{(\alpha - 1)^4(\alpha - 2)(\alpha - 3)(\alpha - 4)} \\
&= \frac{3\beta^4\alpha(3\alpha^2 + \alpha + 2)}{(\alpha - 1)^4(\alpha - 2)(\alpha - 3)(\alpha - 4)}.
\end{aligned}$$

**Lause 1.3.** *Pareto jaotuse asümmeetriakordaja on kujul*

$$\beta_1(X) = \frac{2(\alpha + 1)}{\alpha - 3} \sqrt{\frac{\alpha - 2}{\alpha}}.$$

**Tõestus.** Leiame  $\beta_1(X)$  avaldise tsentraalsete momentide kaudu

$$\beta_1(X) = \frac{E(X - EX)^3}{DX^{\frac{3}{2}}}. \quad (9)$$

Võrduse (9) nimetaja saame dispersiooni avaldisest (4)

$$DX^{\frac{3}{2}} = \left( \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)} \right)^{\frac{3}{2}} = \frac{\alpha^{\frac{3}{2}}\beta^3}{(\alpha-1)^3(\alpha-2)^{\frac{3}{2}}}.$$

Kasutades Pareto jaotuse kolmandat tsentraalset momenti (7) ning asendades leitud väärtused valemisse (9) saamegi Pareto jaotuse asümmeetriakordaja lõpliku avaldise

$$\begin{aligned} \beta_1(X) &= \frac{2\beta^3\alpha(\alpha+1)(\alpha-1)^3(\alpha-2)^{\frac{3}{2}}}{(\alpha-1)^3(\alpha-2)(\alpha-3)\alpha^{\frac{3}{2}}\beta^3} = \\ &= \frac{2\alpha^{-\frac{1}{2}}(\alpha+1)(\alpha-2)^{\frac{1}{2}}}{(\alpha-3)} = \frac{2(\alpha+1)}{\alpha-3} \sqrt{\frac{\alpha-2}{\alpha}}. \end{aligned}$$

**Lause 1.4.** *Pareto jaotuse järsakuskordaja on kujul*

$$\beta_2(X) = \frac{3(3\alpha^3 - 5\alpha^2 - 4)}{\alpha(\alpha-3)(\alpha-4)}.$$

**Tõestus.** Leiame  $\beta_2(X)$  avaldise tsentraalsete momentide kaudu

$$\beta_2(X) = \frac{E(X - EX)^4}{DX^2}. \quad (10)$$

Avaldise (10) nimetaja saame dispersiooni valemist (4)

$$DX^2 = \left( \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)} \right)^2 = \frac{\alpha^2\beta^4}{(\alpha-1)^4(\alpha-2)^2}.$$

Kasutades Pareto jaotuse neljandat tsentraalset momenti (8) ning asendades leitud väärtused valemisse (10) saamegi Pareto jaotuse järsakuskordaja

$$\begin{aligned} \beta_2(X) &= \frac{3\beta^4\alpha(3\alpha^2 + \alpha + 2)(\alpha-1)^4(\alpha-2)^2}{(\alpha-1)^4(\alpha-2)(\alpha-3)(\alpha-4)\alpha^2\beta^4} = \\ &= \frac{3(3\alpha^3 - 5\alpha^2 - 4)}{\alpha(\alpha-3)(\alpha-4)}. \end{aligned}$$

## 1.5 Weibulli jaotus

**Definitsioon. 1.5** Öeldakse, et juhuslik suurus  $X$  on Weibulli jaotusest kujuparameetriga  $\alpha$  ja skaalaparameetriga  $\lambda$ ,  $X \sim W(\alpha, \lambda)$ , kus  $\alpha > 0$  ja  $\lambda > 0$ , kui tema tihedusfunktsioon avaldub kujul

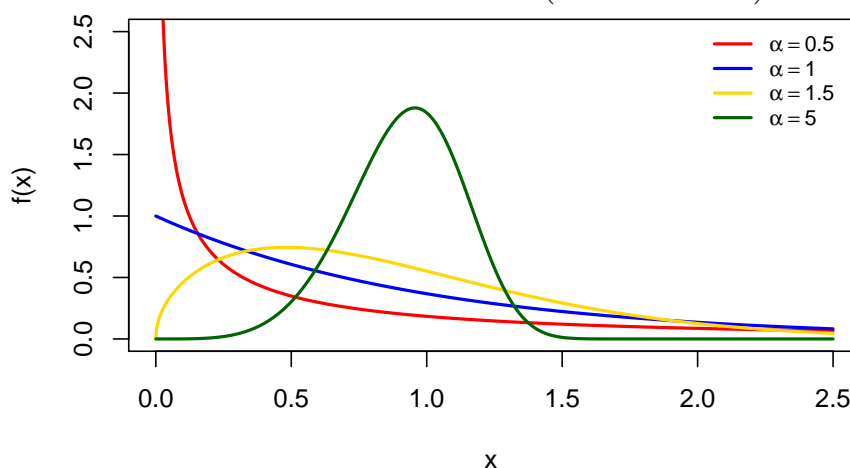
$$f_X(x) = \frac{\alpha}{\lambda} \left(\frac{x}{\lambda}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\lambda}\right)^\alpha\right], \quad x \geq 0.$$

Weibulli jaotuse jaotusfunktsioon avaldub järgmisel kujul

$$F_X(x) = 1 - \exp\left[-\left(\frac{x}{\lambda}\right)^\alpha\right].$$

Weibulli jaotuse juhusliku suuruse  $X$  arvarakteristikud on järgmised:

- keskväärtus:  $EX = \lambda\Gamma\left[\frac{1+\alpha}{\alpha}\right]$ ;
- dispersioon:  $DX = \lambda^2[\Gamma\left[\frac{2+\alpha}{\alpha}\right] - (\Gamma\left[\frac{1+\alpha}{\alpha}\right])^2]$ ;
- asümeetriakordaja:  $\beta_1(X) = \frac{\Gamma\left[\frac{3+\alpha}{\alpha}\right] - 3\Gamma\left[\frac{1+\alpha}{\alpha}\right]\Gamma\left[\frac{2+\alpha}{\alpha}\right] + 2(\Gamma\left[\frac{1+\alpha}{\alpha}\right])^3}{(\Gamma\left[\frac{2+\alpha}{\alpha}\right] - (\Gamma\left[\frac{1+\alpha}{\alpha}\right])^2)^{\frac{3}{2}}}$ ;
- järsakuskordaja:  $\beta_2(X) = \frac{\Gamma\left[\frac{4+\alpha}{\alpha}\right] - 4\Gamma\left[\frac{3+\alpha}{\alpha}\right]\Gamma\left[\frac{1+\alpha}{\alpha}\right] + 6\Gamma\left[\frac{2+\alpha}{\alpha}\right](\Gamma\left[\frac{1+\alpha}{\alpha}\right])^2 - 3(\Gamma\left[\frac{1+\alpha}{\alpha}\right])^4}{(\Gamma\left[\frac{2+\alpha}{\alpha}\right] - (\Gamma\left[\frac{1+\alpha}{\alpha}\right])^2)^2}$ .



Joonis 5: Weibulli jaotuse tihedusfunktsioon, kui  $\lambda = 1$ .

**Lause 1.5.** Weibulli jaotuse  $k$ -s moment avaldub kujul

$$EX^k = \lambda^k \Gamma \left[ \frac{k + \alpha}{\alpha} \right].$$

**Tõestus.** Näitame, et antud võrdus kehtib

$$\begin{aligned} EX^k &= \int_0^\infty x^k f(x) dx = \int_0^\infty x^k \frac{\alpha}{\lambda} \left( \frac{x}{\lambda} \right)^{\alpha-1} \exp \left[ - \left( \frac{x}{\lambda} \right)^\alpha \right] dx \\ &= \alpha \int_0^\infty x^{k-1} \left( \frac{x}{\lambda} \right)^\alpha \exp \left[ - \left( \frac{x}{\lambda} \right)^\alpha \right] dx. \end{aligned}$$

Tulemuse saamiseks teeme muutuja vahetuse  $t = \left( \frac{x}{\lambda} \right)^\alpha$ , millest järeldub, et  $x = \lambda t^{\frac{1}{\alpha}}$  ning sellest omakorda saame, et  $dx = \frac{1}{\alpha} \lambda t^{\frac{1}{\alpha}-1} dt$ .

$$\begin{aligned} EX^k &= \alpha \int_0^\infty (\lambda t^{\frac{1}{\alpha}})^{k-1} t \exp[-t] \frac{1}{\alpha} \lambda t^{\frac{1}{\alpha}-1} dt \\ &= \lambda^k \int_0^\infty t^{\frac{k-1}{\alpha}} t^{\frac{1}{\alpha}-1} t \exp[-t] dt \\ &= \lambda^k \int_0^\infty t^{\frac{k}{\alpha}} \exp[-t] dt = \lambda^k \int_0^\infty t^{\frac{k}{\alpha}+1-1} \exp[-t] dt \\ &= \lambda^k \int_0^\infty t^{(\frac{k}{\alpha}+1)-1} \exp[-t] dt = \lambda^k \int_0^\infty t^{(\frac{k+\alpha}{\alpha})-1} \exp[-t] dt \\ &= \lambda^k \Gamma \left[ \frac{k + \alpha}{\alpha} \right]. \end{aligned}$$

Viimases integraalis tunneme ära gammafunktsiooni kuju (2), mille parameeter on  $\frac{k+\alpha}{\alpha}$ .

Seega Weibulli jaotuse esimesed neli momenti on järgmised

$$\begin{aligned} EX &= \lambda \Gamma \left[ \frac{1 + \alpha}{\alpha} \right]; \\ EX^2 &= \lambda^2 \Gamma \left[ \frac{2 + \alpha}{\alpha} \right]; \\ EX^3 &= \lambda^3 \Gamma \left[ \frac{3 + \alpha}{\alpha} \right]; \\ EX^4 &= \lambda^4 \Gamma \left[ \frac{4 + \alpha}{\alpha} \right]. \end{aligned}$$

**Lause 1.6.** Weibulli jaotuse asümmeetriakordaja avaldub momentide kaudu kujul

$$\beta_1(X) = \frac{E(X^3) - 3EXE(X^2) + 2(EX)^3}{DX^{\frac{3}{2}}}.$$

**Tõestus:** Avame sulud  $\beta_1(X)$  avaldises

$$\begin{aligned} \beta_1(X) &= \frac{E(X - EX)^3}{DX^{\frac{3}{2}}} \\ &= \frac{E(X^3 - 3X^2EX + 3X(EX)^2 - (EX)^3)}{DX^{\frac{3}{2}}} \\ &= \frac{E(X^3) - 3EXE(X^2) + 2(EX)^3}{DX^{\frac{3}{2}}}. \end{aligned}$$

Pärast momentide avaldiste asendamist  $\beta_1(X)$  avaldisse saamegi asümmeetriakordaja gammafunktsioonide kaudu eespool toodud kujul.

**Lause 1.7.** Weibulli jaotuse järsakuskordaja avaldub momentide kaudu kujul

$$\beta_2(X) = \frac{E(X^4) - 4E(X^3)EX + 6E(X^2)(EX)^2 - 3(EX)^4}{DX^2}.$$

**Tõestus:** Avame sulud  $\beta_2(X)$  avaldises

$$\begin{aligned}\beta_2(X) &= \frac{E(X - EX)^4}{DX^2} \\ &= \frac{E(X^4 - 4X^3EX + 6X^2(EX)^2 - 4X(EX)^3 + (EX)^4)}{DX^2} \\ &= \frac{E(X^4) - 4E(X^3)EX + 6E(X^2)(EX)^2 - 3(EX)^4}{DX^2}.\end{aligned}$$

Pärast momentide avaldiste asendamist  $\beta_2(X)$  avaldisse saamegi Weibulli jaotuse järsakuskordaja gammafunktsioonide kaudu eespool toodud kujul.

## 2 Jääkeskmise meetod (*Mean excess method*)

Käesolevas peatükis anname ülevaate jääkeskmise meetodist. Tuletame üldise valemi jääkeskmise leidmiseks ning valemid Pareto ja log-normaalse jaotuse korral. Tulemused on esitatud raamatute (Hogg, Klugmann, 1984; Hall, Wellner 1981) põhjal.

Ekstreemselt suured kahjusummad võivad mõjutada kindlustusettevõtte edasist toimimist. Seega tuleb erilist tähelepanu pöörata just kahjudele, mis on toonud endaga kaasa suure kahjusumma. Analüüsida tuleb kahjude jaotuse sabaosa ning leida, milline jaotus sobitub andmetele kõige paremini. Edaspidi kahjusid modelleerides on vaja hinnata võimalikult täpselt tõenäosust teatud suurust ületavate kahjude esinemiseks.

### 2.1 Meetodi tõenäosuslik kirjeldus

**Definitsioon. 2.1.** Olgu juhuslik suurus  $X$  tõenäosustihedusega  $f(x)$  ning  $x > 0$ , siis juhusliku suuruse  $X$  jääkeskmine punktis  $x$  on määratud valemiga

$$e(x) = \int_x^\infty (w - x) \frac{f(w)}{P(X \geq x)} dw. \quad (11)$$

Kui juhusliku suuruse  $X$  jaotusfunktsioon on  $F(x) = P(X \leq x)$ , siis kasutades ositi integreerimise valemit  $\int u dv = uv - \int v du$  ning valides  $u = (x - w)$  ja  $dv = f(x) dw$  saame, et  $du = -dw$  ning  $v = 1 - F(w)$ . Kasutades saadud tulemust, saame leida

jääkkeskmisele alternatiivse kuju

$$\begin{aligned}
 e(x) &= \int_x^\infty (w-x) \frac{f(w)}{P(X \geq x)} dw = \frac{1}{P(X \geq x)} \int_x^\infty (w-x) f(w) dw \\
 &= \frac{1}{1-F(x)} \left( [(x-w)[1-F(w)]]_x^\infty + \int_x^\infty [1-F(w)] dw \right) \\
 &= \frac{\int_x^\infty [1-F(w)] dw}{1-F(x)} \tag{12}
 \end{aligned}$$

eeldusel, et

$$\lim_{b \rightarrow \infty} (x-b)[1-F(b)] = 0.$$

Kasutades valemeid (11) ja (12) on võimalik leida jääkkeskmise valemid meie poolt kasutatud jaotustele.

Jääkkeskmise  $e(x)$  hinnanguks on empiiriline jääkkeskmise kujul

$$e_n(x) = \frac{\sum_{i=0}^n I_{\{x_i \geq x\}} x_i}{\#\{x_i \geq x\}} - x.$$

Näeme, et empiiriline jääkkeskmise on lihtsasti leitav. Me jagame kõigi kahjude  $x_i$  summa, kus  $x_i \geq x$ , nende arvuga, millest lahutame suuruse  $x$ . Teades erinevate jaotuste jääkkeskmise valemid ning nende kujusid on võimalik võrdluse teel saada esialgset informatsiooni, milline jaotus võiks sobituda antud andmete sabaosale kõige paremini.

## 2.2 Pareto ja log-normaalse jaotuse jääkkeskmised

Edasise analüüsi käigus selgus, et kõige paremini sobivateks jaotusteks on log-normaalne jaotus ning kahjuandmete sabaosal Pareto jaotus. Leiame järgnevalt jääkkeskmise avaldised nende jaotuste korral.

**Lause 2.1.** Pareto jaotuse jääkkeskmise  $e(x)$  on kujul

$$e(x) = \frac{\beta + x}{\alpha - 1}.$$

**Tõestus.** Selleks, et Pareto jaotuse korral oleks võimalik leida keskväärtust peab kehtima  $\alpha > 1$ . Kasutades Pareto jaotuse jaotusfunktsiooni (3) ning jääkkeskmise valemit (12) saame leida Pareto jääkkeskmise

$$\begin{aligned} e(x) &= \frac{\int_x^\infty [1 - F(w)]dw}{1 - F(x)} = \frac{\int_x^\infty 1 - 1 + \left(\frac{\beta}{\beta+w}\right)^\alpha dw}{1 - 1 + \left(\frac{\beta}{\beta+x}\right)^\alpha} \\ &= \frac{\int_x^\infty \left(\frac{\beta}{\beta+w}\right)^\alpha dw}{\left(\frac{\beta}{\beta+x}\right)^\alpha} = \frac{\beta^\alpha \int_x^\infty \frac{1}{(\beta+w)^\alpha} dw}{\left(\frac{\beta}{\beta+x}\right)^\alpha}. \end{aligned}$$

Vaatame eraldi integraali

$$\begin{aligned} \int_x^\infty \frac{1}{(\beta+w)^\alpha} dw &= \int_x^\infty (\beta+w)^{-\alpha} dw = \left. \frac{(\beta+w)^{-\alpha+1}}{-\alpha+1} \right|_x^\infty \\ &= \frac{(\beta+\infty)^{1-\alpha}}{1-\alpha} - \frac{(\beta+x)^{1-\alpha}}{1-\alpha} \\ &= 0 + \frac{(\beta+x)^{1-\alpha}}{\alpha-1} \\ &= \frac{(\beta+x)^{1-\alpha}}{\alpha-1}. \end{aligned}$$

Asendades saadud avaldise eespool leitud valemisse saame

$$\begin{aligned} e(x) &= \frac{\beta^\alpha \int_x^\infty \frac{1}{(\beta+w)^\alpha} dw}{\left(\frac{\beta}{\beta+x}\right)^\alpha} = \frac{\beta^\alpha (\beta+x)^{1-\alpha} (\beta+x)^\alpha}{(\alpha-1)\beta^\alpha} \\ &= \frac{\beta+x}{\alpha-1}. \end{aligned}$$

**Lause 2.2.** Log-normaalse jaotuse jääkeskmise  $e(x)$  on kujul

$$e(x) = \frac{e^{\frac{\sigma^2}{2} + \mu} \left[ 1 - \Phi \left( \frac{\ln x - \mu - \sigma^2}{\sigma} \right) \right]}{1 - \Phi \left( \frac{\ln x - \mu}{\sigma} \right)} - x.$$

**Tõestus.** Näitame, et ülaltoodud valem kehtib

$$e(x) = \int_x^\infty (w - x) \frac{f(w)}{P(X \geq x)} dw = \frac{1}{1 - F(x)} \int_x^\infty wf(w) dw - x,$$

kus log-normaalse jaotuse tihedus on

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

Järgmisena leiame integraali väärtuse, selleks teeme muutujavahetuse

$$y = \frac{\ln w - \mu}{\sigma} \Rightarrow (\sigma y + \mu) = \ln w \Rightarrow w = e^{\sigma y + \mu};$$

$$dy = \frac{1}{w\sigma} dw \Rightarrow dw = w\sigma dy.$$

Koos muutuja vahetusega teiseneb ka alumine raja. Kui  $w = x$ , siis

$$y = \frac{\ln x - \mu}{\sigma} = z_0.$$

Leiame integraali

$$\begin{aligned} & \int_x^\infty \frac{w}{\sqrt{2\pi}\sigma w} e^{-\frac{1}{2\sigma^2}(\ln w - \mu)^2} dw \\ &= \int_{z_0}^\infty \frac{1}{\sqrt{2\pi}\sigma} w\sigma e^{-\frac{1}{2}y^2} dy \\ &= \int_{z_0}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y^2 - 2\sigma y - 2\mu)} dy. \end{aligned}$$

Teisendame  $e$  astendajat

$$-\frac{1}{2}(y^2 - 2\sigma y + \sigma^2 - \sigma^2 - 2\mu) = -\frac{1}{2}(y - \sigma)^2 + \frac{\sigma^2}{2} + \mu.$$

Seega

$$\begin{aligned} e^{\frac{\sigma^2}{2} + \mu} \int_{z_0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y - \sigma)^2} dy &= e^{\frac{\sigma^2}{2} + \mu} F_{N(\sigma, 1)}(y) \Big|_{z_0}^{\infty} \\ &= e^{\frac{\sigma^2}{2} + \mu} [1 - F_{N(\sigma, 1)}(z_0)]. \end{aligned}$$

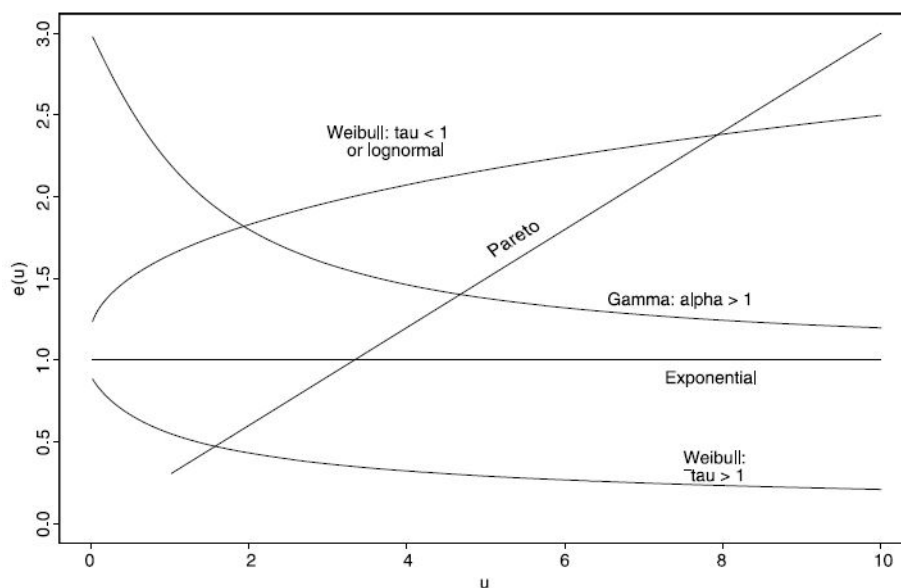
Kui  $X \sim N(\sigma, 1)$ , siis  $Z = X - \sigma \sim N(0, 1)$  ning saame, et

$$\begin{aligned} F_{N(\sigma, 1)}(z_0) &= P(X_{N(\sigma, 1)} \leq z_0) \\ &= P(X - \sigma \leq z_0 - \sigma) = \Phi(z_0 - \sigma). \end{aligned}$$

Pannes tulemused kokku, saame

$$\begin{aligned} e(x) &= \frac{e^{\frac{\sigma^2}{2} + \mu} \left[ 1 - \Phi\left(\frac{\ln x - \mu}{\sigma} - \sigma\right) \right]}{1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)} - x \\ &= \frac{e^{\frac{\sigma^2}{2} + \mu} \left[ 1 - \Phi\left(\frac{\ln x - \mu - \sigma^2}{\sigma}\right) \right]}{1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)} - x. \end{aligned}$$

Lugedes kirjandust jääkkeskmise meetodi kohta, on raske leida erinevate jaotuste täpseid lõppvalemuid. Tihtilugu on antud ligikaudsed valemid, mis sisaldavad konstante ning konstantide kuju ei ole eraldi välja toodud. Küll aga võib leida erinevatest allikatest graafikud tuntumate jaotuste jääkkeskmiste kujude kohta ning kuidas kujud sõltuvad antud jaotuse parameetritest. Toome järgnevalt ära raamatu (Embrecht *et al.*, 2008, lk 295), põhjal joonise, kus on esitatud jääkkeskmiste kujud erinevate jaotuste korral.



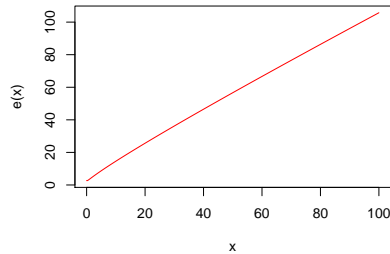
Joonis 6: Jääkkeskmiste graafikud.

Lisaks on raamatus (Embrecht *et al.*, 2008, lk 161) ära toodud nende jaotuste jääkkeskmiste avaldised kujul

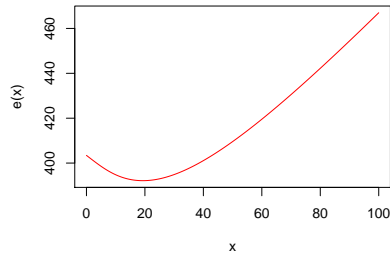
- Pareto  $e(u) = \frac{\kappa+u}{\alpha-1}$ ,  $\alpha > 1$ ;
- Log-normaalne  $e(u) = \frac{\sigma^2 u}{\ln u - \mu} (1 + o(1))$ ;
- Weibulli  $e(u) = \frac{u^{1-\tau}}{c\tau} (1 + o(1))$ ;
- Gamma  $e(u) = \beta^{-1} \left( 1 + \frac{\alpha-1}{\beta u} + o\left(\frac{1}{u}\right) \right)$ .

Üks olulisemaid jaotusi, mida kasutatakse kahjude modelleerimisel on log-normaalne jaotus. Tema kuju on joonisel (6) esitatud ühe parameetrite komplekti korral ning ei ole teada, kuidas see kuju sõltub parameetritest. Magistritöö käigus leidis autor, et log-normaalne jaotuse jääkkeskmise kuju muutub sõltuvalt parameetritest küllaltki palju ning järgnevalt toome siinkohal välja log-normaalne jaotuse jääkkeskmiste

kujud mõnede erinevate parameetrikombinatsioonide korral.

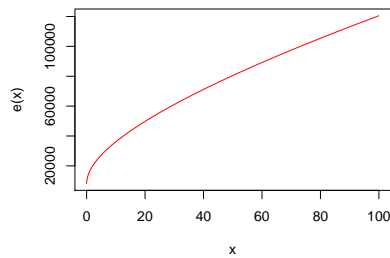


(a)  $\mu = 0$

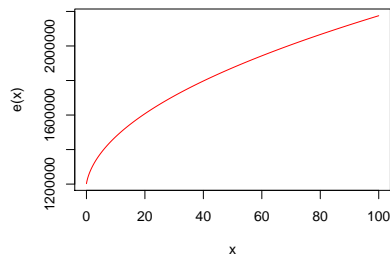


(b)  $\mu = 5$

Joonis 7: Log-normaalse jaotuse jääkeskmise kujud erineva  $\mu$  korral, kui  $\sigma = 1$ .

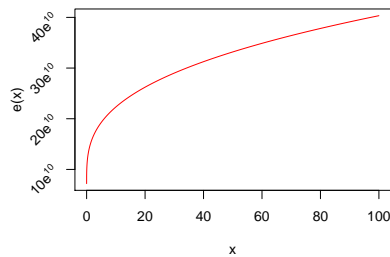


(a)  $\mu = 0$

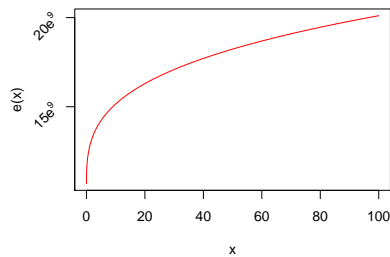


(b)  $\mu = 5$

Joonis 8: Log-normaalse jaotuse jääkeskmise kujud erineva  $\mu$  korral, kui  $\sigma = 3$ .



(a)  $\mu = 0$



(b)  $\mu = 5$

Joonis 9: Log-normaalse jaotuse jääkeskmise kujud erineva  $\mu$  korral, kui  $\sigma = 5$ .

## 3 Kahjude analüüs

### 3.1 Andmete kirjeldus

Järgnevalt anname ülevaate töö praktilises osas kasutatud andmetest. Töös kasutame kindlustusseltsi If P&C Insurance AS kaskokahjude andmeid. Originaalandmestikus on andmed kolme Balti riigi kaskokahjude kohta, antud töös võtame vaatluse alla Eesti andmed. Lõplikus andmestikus on 117 000 rida, kus igäüks esindab eraldiseisvat kaskokahjut. Andmestikus on 7 tunnust, nendeks on:

- **Country** - näitab, mis riigi kaskokahjuga on tegemist (antud töö lõplikus andmestikus vaid koodiga *EE* kahjud)
- **EventDate** - näitab kahju toimumise kuupäeva
- **ReportingMonth** - näitab, mis aastal ja kuul on kahjust teavitatud kindlustusseltsi
- **RiskEng** - näitab, mis liiki kahjuga on tegemist (tunnus inglise keeles)
- **UltimateIncurred** - näitab välja makstud või potentsiaalselt välja makstava hüvitise suurust (maha on arvatud omavastutus)
- **ClientType** - näitab, kas tegemist on erakliendi või juriidilise isikuga
- **VehicleAge** - näitab, kui vana oli sõiduk täisaastates kahju tekkimise hetkel.

Töös kasutatavate kahjuandmete puhul on kahjusummast omavastutus juba maha arvatud, seetõttu on väga väikeste kahjude korral tegemist reeglina omavastutust ületava summaga. Andmestikku on kaasatud ainult kahjud, mis olid suuremad või võrdsed 1 euroga. Kahjude toimumise kuupäevad jäävad vahemikku 23.06.2008 - 31.12.2020 ning nendest teatamine vahemikku jaanuar 2010 - detsember 2020. Andmestikus on kokku 39 eri liiki kahjut, nendest kolm suurima sagedusega kahjutüüpi on järgmised:

- **Breaking or damaging (windscreen)** - esiklaasi purunemine või vigastamine. Kokku 45 799 juhtumit ehk ligikaudu 39,1% kõikidest kahjudest.
- **Hitting other object** - muu objekti tabamine või sellele otsasõit. Kokku 21 878 juhtumit ehk ligikaudu 18,7% kõikidest kahjudest.
- **Damaging parked vehicle** - parkiva sõiduki kahjustamine. Kokku 16 506 juhtumit ehk ligikaudu 14,1% kõikidest kahjudest.

Näeme, et kolm kõige sagedasemat kahjukategooriat moodustavad kokku 71,9% kõikidest kahjudest. Antud 117 000 kahjust on tekitatud 77 960 ehk 66,6% eraklientide poolt ning 39 040 ehk 33,4% juriidiliste isikute poolt. Sõiduki vanuse varieeruvus on suur - vanim sõiduk, millega kahju on tekitatud, oli 48-aastane. Teiste kahjude puhul jääb sõiduki vanus täisaastates vahemikku 0 – 25 aastat. Suurem osa kahjudest on tekitatud uute sõidukitega, mille vanus jääb vahemikku 0 – 6 aastat. Nende sõidukitega on kokku tekitatud 84 027 kahju, mis moodustavad 71,8% kõikidest kahjudest.

Tabelis 1 on toodud kaskokahjude hüvitiste peamised empiirilised arvarakteristikud. Tabelist näeme, et minimaalne hüvitise suurus on 1,03 eurot ning maksimaalne hüvitise suurus 69 454,42 eurot. Keskmise kahju antud andmete puhul on 959,22 eurot ja mediaankahju 474,08 eurot. Antud kahjuandmete standardhälve on 1720,12 eurot, seega hüvitiste suurused varieeruvad palju. Keskmise kahju on veidi üle kahe korra suurem kui mediaankahju, mida kinnitab ka fakt, et asümmeetrijakordaja on positiivne ning omab väärtust 10,27. Järsakuskordaja antud kahjuandmete puhul on 215,05, mis näitab, et tegemist on raske sabaga jaotusega.

Tabel 1: Kaskokahjude hüvitiste empiirilised arvarakteristikud.

	<b>Hüvitise suurus</b>
<b>Miinumum</b>	1,03
<b>Maksimum</b>	69454,42
<b>Keskmine</b>	959,22
<b>Mediaan</b>	474,08
<b>Standardhälve</b>	1720,12
<b>Asümmetriakordaja</b>	10,27
<b>Järsakuskordaja</b>	215,05

Tabelis 2 on toodud kaskokahjude hüvitiste kvantiilid. Näeme, et antud kahjude 95% kvantiil on 3231,43 eurot, mis tähendab seda, et 95% kõikidest kahjudest on väiksemad kui 3231,43 eurot. Kõikide kaskokahjude hüvitiste summa kokku on 112 228 867 eurot. Antud kahjude puhul on 5% suurimate kahjude kogusumma 36 673 198 eurot, mis moodustab kogukahjust ligikaudu 33%. Seega on eriti tähtis, et hilisema sobivuse testimisel annaks sobitatud jaotus adekvaatse hinnangu suurimate kahjude suurustele jaotuse sabaosal.

Tabel 2: Kaskokahjude hüvitiste kvantiilid.

<b>Kvantiil</b>	<b>Hüvitise suurus</b>	<b>Kvantiil</b>	<b>Hüvitise suurus</b>
<b>0,1%</b>	19,05	<b>55%</b>	534,25
<b>0,5%</b>	37,00	<b>60%</b>	606,45
<b>1%</b>	50,00	<b>65%</b>	702,23
<b>5%</b>	126,91	<b>70%</b>	822,10
<b>10%</b>	191,10	<b>75%</b>	982,63
<b>15%</b>	224,03	<b>80%</b>	1 189,92
<b>20%</b>	251,88	<b>85%</b>	1 509,82
<b>25%</b>	280,00	<b>90%</b>	2 074,38
<b>30%</b>	310,64	<b>95%</b>	3 231,43
<b>35%</b>	344,30	<b>99%</b>	7 659,52
<b>40%</b>	380,24	<b>99,5%</b>	10 379,90
<b>45%</b>	424,00	<b>99,9%</b>	19 936,49
<b>50%</b>	474,08	<b>100%</b>	69 454,42

## 3.2 Jaotuste sobitamine kahjuandmetele

Järgnevalt sobitame töös kasutatavaid jaotusi kahjuandmetele. Sobitamise käigus soovime esmalt leida jaotuse, mis kirjeldab antud kahjuandmeid kõige täpsemini kogu andmestikul. Teiseks püüame leida punkti, kus jagades kahjuandmed kaheks ning sobitades jaotusi eraldi osadel, saaksime parema sobivuse kahjuandmetega.

Toome välja sobitatud jaotuste parameetrite hinnangud, kahjuandmetele sobitatud jaotuste graafikud ja Kolmogorov-Smirnovi (edaspidi K-S) teststatistiku  $K_n$  ning  $\chi^2$ -statistiku väärtused erinevate jaotuste korral ja kvantiil-kvantiil graafikud.

Kolmogorov-Smirnovi statistik avaldub kujul (nt. Kollo, 2004, lk 29):

$$K_n = \max(K_n^+, K_n^-),$$

kus

- $K_n^+ = \sup_{-\infty < x < \infty} (F_n(x) - F_X(x));$
- $K_n^- = \sup_{-\infty < x < \infty} (F_X(x) - F_n(x));$
- $F_X(x)$  on lähendava jaotuse jaotusfunktsioon;
- $F_n(x) = \frac{\#(x_i | x_i \leq x)}{n}$  on empiiriline jaotusfunktsioon.

Töös kasutatud K-S statistikute väärtuste leidmiseks on kirjutatud rakendustarkvaras R programm, mis etteantud andmestiku põhjal leiab K-S teststatistiku väärtuse. Programmile saab ette anda kahjuandmestikku lähendava jaotuse nimetuse ning osakaalu, kui suure osa andmete pealt hinnatakse jaotuse parameetrid ja kui suure osa peal testitakse jaotuse sobivust. Töös on osakaaluks valitud 0,5 ehk juhusliku valikuga võetakse pool andmestikust jaotuste parameetrite hindamiseks ja pool andmestikku jääb leitud jaotuse sobivuse testimiseks. Teststatistiku väärtused

leitakse kõigile töös kasutatavatele jaotustele ning et tulemused oleks stabiilsemad, on suuremate andmemahude puhul korratud testi 50 korda ning väiksemate andmemahude puhul 100 korda. Lõplikuks tulemuseks on võetud vastavate korduste aritmeetiline keskmine. Lisades on toodud ära tabelid 19 - 28, kus on toodud täpsemalt välja K-S teststatistikute varieeruvus erinevate andmestike ning jaotuspunktide korral.

Teise võimalusena kasutame jaotuste sobivuse kontrollimiseks  $\chi^2$ -testi. Testimaks hüpoteesipaari

$$H_0 : F(x) = F_n(x)$$

$$H_1 : F(x) \neq F_n(x)$$

vaatleme valimit  $x_1, x_2, \dots, x_n$  ja leiame sagedused  $k$  klassis  $(c_0, c_1], \dots, (c_{k-1}, c_k]$  ning tähistame nad vastavalt  $n_1, n_2, \dots, n_k$ . Klassidesse langemise tõenäosused saame leida järgnevalt  $p_i = F(c_i) - F(c_{i-1}), i = 1, 2, \dots, k$  ning oodatavad väärtused klassides on vastavalt  $np_1, np_2, \dots, np_k$ . Olles leidnud eelnevalt mainitud väärtused, saamegi arvutada  $\chi^2$ -statistiku väärtuse (Hogg, Klugman, 1984, lk 103-104):

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (13)$$

Valemis (13) leitud suurus on statistiku  $T_{\chi^2}$  väärtus

$$T_{\chi^2} = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i},$$

kus  $N_i \sim (n, p_i); \sum_{i=1}^k N_i = n$ . Arvu  $n$  kasvades läheneb statistiku  $T_{\chi^2}$  jaotus  $\chi^2$ -jaotusele vabadusastmete arvuga  $(k - 1)$ . Otsuse langetamiseks hüpoteeside kont-

rolli ülesandes arvutame statistiku väärtuse ning fikseerime usaldusnivoo  $\alpha$ . Kui  $\chi^2 \geq h_{k-1;1-\alpha}$ , siis võtame vastu sisuka hüpoteesi  $H_1$ , vastasel korral jääme  $H_0$  juurde. Siin  $h_{k-1;1-\alpha}$  tähistab  $\chi_{k-1}^2$ -jaotuse  $(1 - \alpha)$ -kvantiili.

Nii  $\chi^2$ -testi kui ka histogrammide kujutamiseks on meil vaja leida vahemike arv. Selleks kasutame Freedman-Diaconise reeglit. (Freedman, Diaconis, 1981, lk 453-454). Selle kohaselt optimaalne histogrammi tulba laius  $h$  avaldub kujul

$$h = 2 \frac{IQR}{\sqrt[3]{n}}, \quad (14)$$

kus

- $IQR = Q_3 - Q_1$  - kvartiilide vahe;
- $n$  - vaatluse all olevate väärtuste koguarv.

Võrdus (14) on saadud empiirilise jaotuse ja teoreetilise tiheduse vahe ruudu integraali minimiseerimisel. Selleks, et leida vajalik vahemike arv  $\chi^2$ -testiks tuleb vaatluse all oleva vahemiku pikkus jagada tulba laiusega  $h$

$$k = \frac{\max_i(x_i) - \min_i(x_i)}{h}.$$

Rakendame edaspidi niiviisi leitud histogrammi tulpade arvu  $\chi^2$ -testi läbiviimisel võrdtõenäoliste klasside korral.

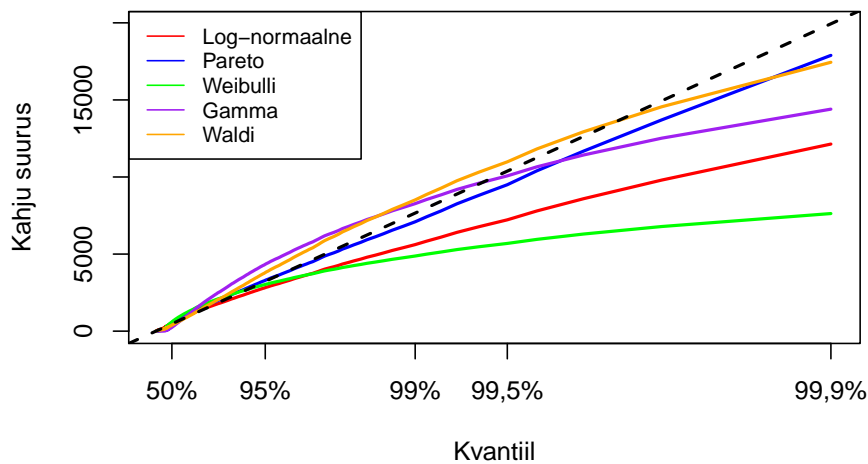
### 3.3 Jaotuste sobitamine kogu kahjuandmestikule

Esmalt sobitame andmetele esimeses peatükis esitatud jaotused, kui vaatluse all on kogu kahjuandmestik. Tabelis 3 on toodud sobitatud jaotuste suurima tõepära meetodil leitud parameetrite hinnangud.

Tabel 3: Parameetrite hinnangud kogu kahjuandmestiku korral.

Jaotus	Parameetrite hinnangud	
<b>Gamma</b>	$\alpha=0,3110$	$\lambda=0,0003$
<b>Waldi</b>	$\mu=959,2211$	$\lambda=301,5103$
<b>Log-normaalne</b>	$\mu=6,2855$	$\sigma=1,0091$
<b>Pareto</b>	$\alpha=2,9027$	$\beta=1825,0630$
<b>Weibulli</b>	$\alpha=0,9048$	$\lambda=901,0458$

Esimene võimalus jaotuste sobivuse hindamiseks on hinnata sobitatud jaotuste tihedusfunktsioone visuaalselt võrreldes andmete empiirilise jaotusega. Ühtlasi võrdleme jaotuste esialgset sobivust kvantiil-kvantiil graafikute abil.

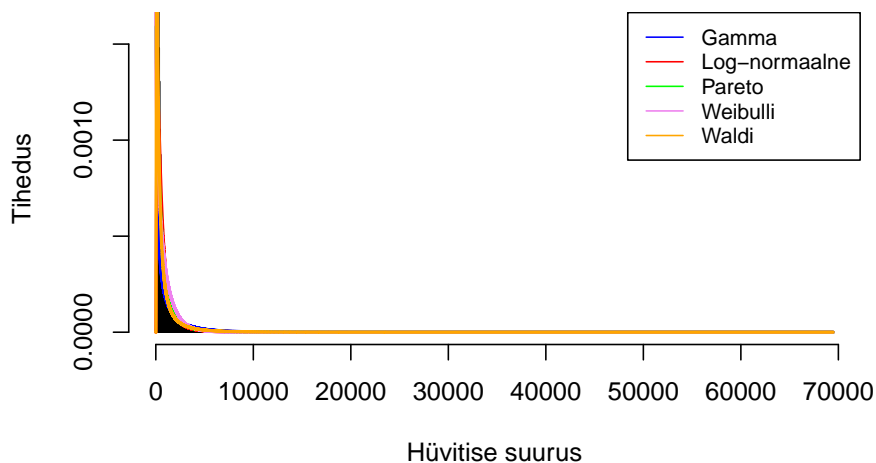


Joonis 10: Kvantiil-kvantiil graafikud kogu kahjuandmestiku korral.

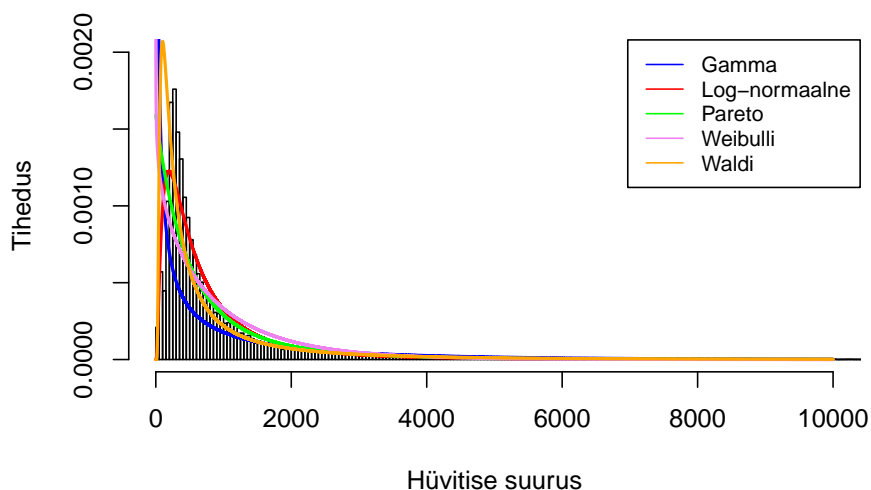
Joonisel (10) on kujutatud kõikide jaotuste kvantiilid võrreldes andmestiku kvantiilidega, mis on joonisel kujutatud musta punktiirjoonena. Jooniselt näeme, et kuni 95% kvantiilini järgivad kogukahjude jaotust log-normaalne, Pareto ning Weibulli jaotus, peale mida Weibulli ja log-normaalne jaotuse väärtused eemalduvad empiirilisest. Suhteliselt punktiirjoone lähedale jääb aga Pareto jaotus. Kuna jooniselt on väga raske välja lugeda kvantiilide ühtivust väiksemate väärtuste korral, siis on lisades välja toodud ka kvantiilide tabel 14 kogukahjude kohta, millest näeme, et log-normaalne jaotus järgib kõige paremini väärtusi, kuni 95%-kvantiilini ning

alates sealt on kõige lähedasemad väärtused Pareto jaotuse korral.

Joonisel (11) on näha kahjude empiirilisele jaotusele sobitatud jaotuste tihedusfunktsioonid. Kuna kahjude kontsentratsioon on kõrge just väikeste väärtuste juures ning antud joonise põhjal ei ole võimalik jaotusi eristada visuaalselt, siis kujutame edaspidi joonistel kahjusid, mis jäävad vahemikku 0 kuni 10 000 eurot. Joonisel (12) on kujutatud nimetatud vahemikku jäävate kahjude empiiriline jaotus ning sobitatud jaotuste tihedusfunktsioonid. Antud joonise põhjal võiks parima sobivuse andmetele anda log-normaalne jaotus, mis tabab suhteliselt hästi ära andmete käitumise just väiksemate kahjude korral. Visuaalsel vaatlusel kõige kehvema sobivuse annab gammajaotus, millele viitab ka kvantiil-kvantiil graafik.



Joonis 11: Kaskokahjude empiiriline jaotus ning sobitatud jaotuste tihedusfunktsioonid.



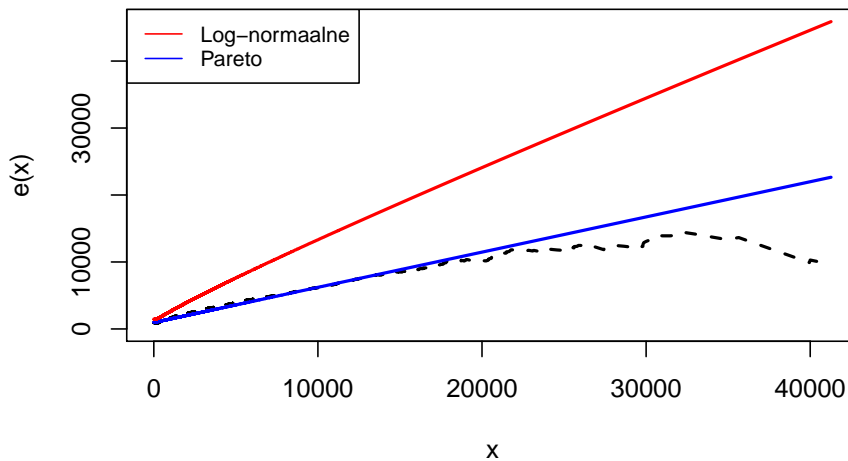
Joonis 12: Kaskokahjude (kuni 10 000 eurot) empiiriline jaotus ning sobitatud jaotuste tihedusfunktsioonid.

Selleks, et saada arvulisi tulemusi jaotuste sobivuse kohta on läbi viidud eespool kirjeldatud K-S test ja  $\chi^2$ -test. Tulemused on toodud tabelis 4. Tabelist näeme, et K-S testi tulemuste puhul on kõige paremini sobivaks jaotuseks log-normaalne jaotus, mille teststatistiku väärtus on  $K_n = 0,0547$  ning kõige halvemini sobiv gammajaotus, mille teststatistiku väärtus on  $K_n = 0,3721$ . Antud testi puhul on kõigi jaotuste teststatistikud suuremad, kui kriitiline väärtus, mis antud andmemahu puhul on 0,00397 ehk K-S testi korral võtame vastu hüpoteesi  $H_1$ , mis väidab jaotuste mitesobivust andmetele. Lisaks on toodud tabelis ka  $\chi^2$ -testi tulemused. Kõigi jaotuste korral on teststatistiku väärtus suurem kriitilisest väärtusest ning tulemuste põhjal on endiselt kõige paremini sobiv jaotus log-normaalne jaotus.

Tabel 4: Sobivusnäitajate väärtused kogu kahjuandmestku korral.

Jaotus	$K_n$	$K_n - \text{kriitiline}$	$\chi^2$	$\chi^2 - \text{kriitiline}$
<b>Gamma</b>	0,3721	0,00397	172360,2	2533,511
<b>Waldi</b>	0,1833	0,00397	80502,03	2533,511
<b>Log-normaalne</b>	0,0547	0,00397	64893,04	2533,511
<b>Pareto</b>	0,1525	0,00397	82826,87	2533,511
<b>Weibulli</b>	0,1224	0,00397	105931,7	2533,511

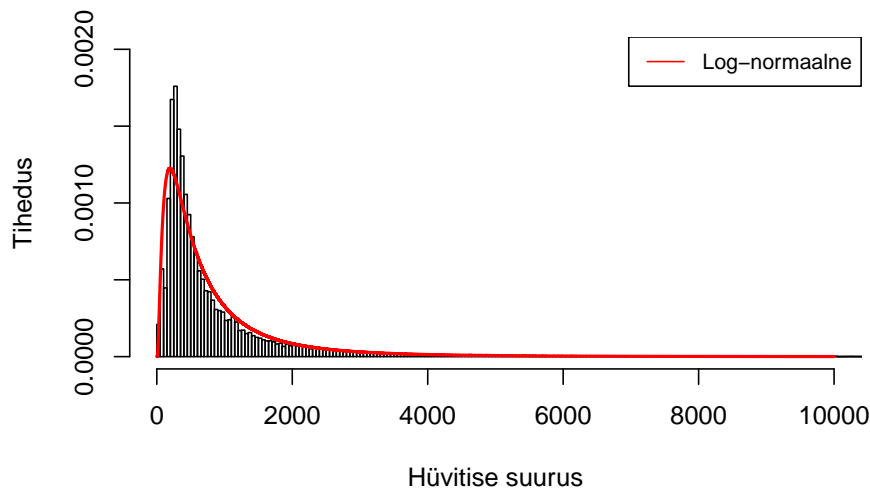
Järgmise võimalusena uurime, milline jaotus võiks andmetele sobitada kasutades jääkkeskmise meetodit. Selle meetodi puhul on teada graafikute kujud erinevate jaotuste korral ning potentsiaalse kandidaadi valimiseks tuleb leida empiiriline jääkkeskmine, kanda see graafikule ning seda võrrelda jaotuste jääkkeskmise graafikutega. Joonisel (13) on kujutatud andmete empiiriline jääkkeskmine musta punktiirjoonega ning log-normaalse ja Pareto jaotuse jääkkeskmise graafikud. Joonise põhjal võiks sobiv kandidaat olla Pareto jaotus, mis ühtib andmete empiirilise jääkkeskmisega suuremas osas andmestikust. Joonis viitab sellele, et kaugemal sabaosal on Pareto jaotus parema sobivusega, kui log-normaalne jaotus.



Joonis 13: Jääkkeskmise meetodi graafikud kogu kahjuandmestikule.

Tuginedes eelpool toodud analüüsile saame väita, et ükski toodud jaotustest ei kirjelda kogukahjusid hästi, kuid parima sobivuse võiks anda siiski log-normaalne jaotus.

tus parameetritega  $\mu = 6,2855$  ning  $\sigma = 1,0091$ . Joonisel (14) on parema ülevaate saamiseks välja toodud kahjude empiiriline jaotus ning sobitatud log-normaalne jaotuse tihedusfunktsioon.



Joonis 14: Kaskokahjude empiiriline jaotus ning sobitatud log-normaalne jaotus.

### 3.4 Andmestiku jagamine kaheks kahjusuuruse kaudu

Eelmises punktis esitatud jääkeskmiste graafikud viitasid sellele, et sabaosal võiks Pareto jaotus olla paremaks lähendiks kui log-normaalne jaotus. Järgnevalt soovime leida kahju suuruse, kust jagada andmestik kaheks ning sobitada jaotused mõlemale osale eraldi. Sellise punkti leidmiseks on koostatud programm, mis jagab andmestiku etteantud kohast kaheks ning sobitab töös vaadeldavaid jaotusi mõlemale osale eraldi. Tabelis 5 on toodud kvantiilid, mille kohalt on andmestik kaheks jagatud, kahju suurused, mis vastavad antud kvantiilidele, K-S teststatistiku väärtused kõige paremini sobitunud jaotuste korral nii andmete esiosale kui ka sabaosale ning leitud teststatistikute väärtuste summa. Otsustuskriteerium parima jaotuspunkti leidmiseks on järgmine. Esiteks peab olema nii andmete esiosale kui ka sabaosale kõige paremini sobitunud jaotuse teststatistik parem, kui seda oli kogu kahjuand-

mestikule sobitunud teststatistik  $K_n = 0.0547$ . Järgmiseks leiame teststatistikute summa ning kahju suuruse, kus see summa on kõige väiksem valimegi jaotuspunktiks. Tabelist 5 näeme, et parimaks punktiks on 95% kvantiil ehk kahju suurus 3231,43 eurot.

Tabel 5: Kolmogorov-Smirnovi teststatistiku väärtused parimatele jaotustele.

Kvantiil	Kahju suurus	$K_n$ – esiosa	$K_n$ – sabaosa	Summa
<b>95%</b>	3231,43	0,0436	0,0258	0,0694
<b>96%</b>	3710,02	0,0445	0,0260	0,0705
<b>97%</b>	4359,21	0,0457	0,0306	0,0763
<b>98%</b>	5394,32	0,0468	0,0296	0,0764
<b>99%</b>	7659,52	0,0493	0,0443	0,0936

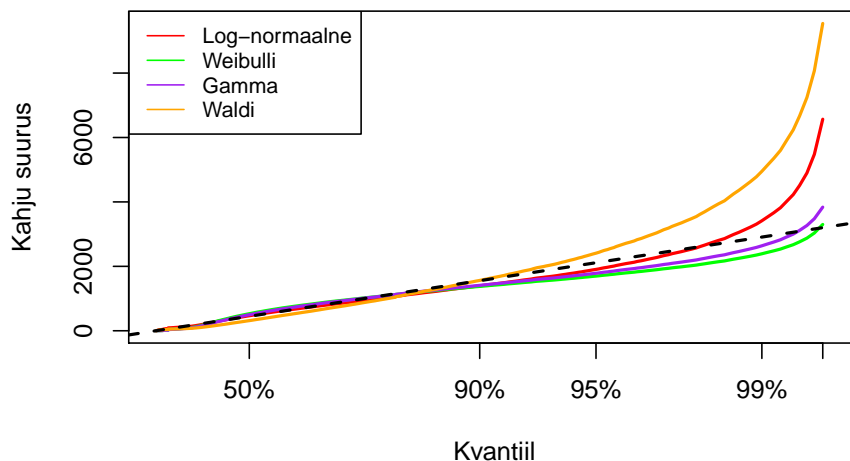
### 3.4.1 Jaotuste sobitamine kahjuandmestiku esiosale

Võtame vaatluse alla kahjuandmete esiosa. Tabelis 6 on toodud sobitatud jaotuste suurima tõepära meetodil leitud parameetrite hinnangud.

Tabel 6: Parameetrite hinnangud kahjuandmestiku esiosal.

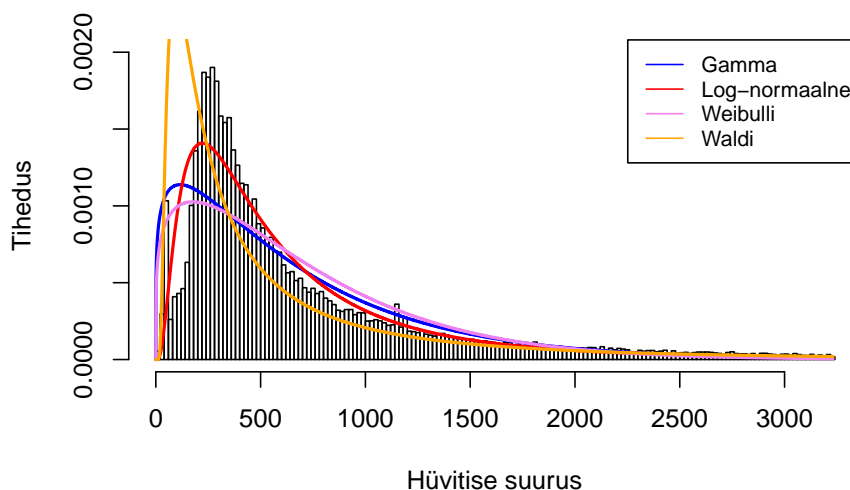
Jaotus	Parameetrite hinnangud	
<b>Gamma</b>	$\alpha=1,2054$	$\lambda=0,0018$
<b>Waldi</b>	$\mu=679,7631$	$\lambda=287,3005$
<b>Log-normaalne</b>	$\mu=6,1634$	$\sigma=0,8732$
<b>Pareto</b>	$\alpha=-9,7371$	$\beta=-7298,6840$
<b>Weibulli</b>	$\alpha=1,2161$	$\lambda=729,9419$

Joonisel (15) on kujutatud jaotuste kvantiilid võrreldes kahjuandmestiku esiosa kvantiilidega. Jooniselt näeme, et log-normaalne, Weibulli ning gammajaotuse kvantiilid on väga lähedased kahjuandmestiku kvantiilidega. Kuna jooniselt on väga raske välja lugeda jaotuste kvantiilide ühtivust kuni 95% kvantiilini, siis on lisades välja toodud ka kvantiilide tabel 15 kahjuandmestiku esiosa kohta. Tabelist näeme, et väiksemate kvantiilide korral ühtib kahjuandmestikuga paremini log-normaalne jaotus ning suuremate kvantiilide puhul annavad paremaid tulemusi Weibulli ja gammajaotus.



Joonis 15: Kvantiil-kvantiil graafikud kahjuandmestiku esiosale.

Joonisel (16) on näha kahjude empiirilisele jaotusele sobitatud jaotuste tihedusfunktsioonid. Kahjuandmestiku esiossa jäävad kahjud suuruses 1 – 3231,43 eurot ning nad moodustavad 95% kõikide kahjude arvust, mis omakorda moodustavad 67% kogukahjust. Antud joonise põhjal annab kõige parema sobivuse log-normaalne jaotus. Lisaks on näha jooniselt, et kahjuandmete esiosaga ei sobitu üldse Waldi jaotus. Andmete esiosa korral ei olnud võimalik defineerida Pareto jaotust kuna jaotuse hinnatud parameetrid tulid negatiivsed. Seda võib seletada sellega, et Pareto jaotuse puhul on tegemist väga raske sabaga jaotusega ja hetkel vaatluse all olevatelt andmetelt oleme raske saba eemaldanud.



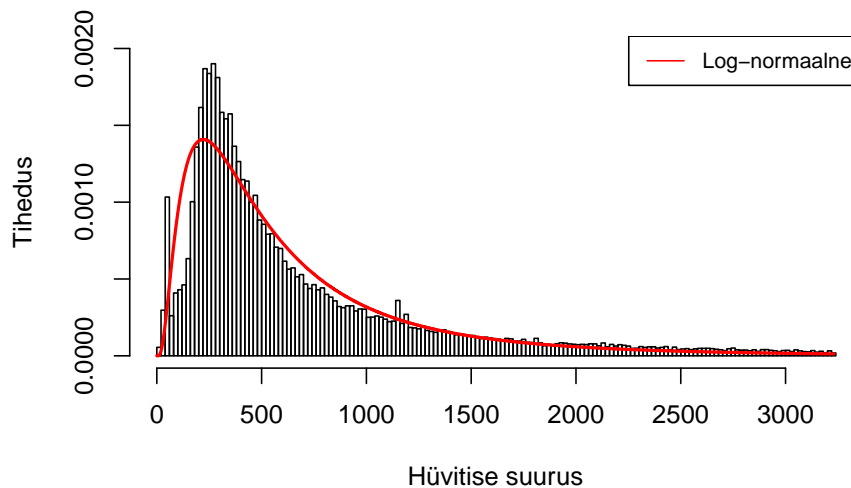
Joonis 16: Kaskokahjude empiiriline jaotus kahjuandmete esiosa korral ning sobitatud jaotuste tihedusfunktsioonid.

Eespool kirjeldatud tulemusi kinnitab ka tabel 7, kus on toodud K-S testi ja  $\chi^2$ -testi tulemused ning nende kriitilised väärtused. Tabeli 7 põhjal annab parima sobivuse kahjuandmete esiosale log-normaalne jaotus, mille K-S teststatistik on  $K_n = 0,0434$ . Kõige vähem sobib kahjuandmete esiosale Waldi jaotus, mille teststatistik väärtus on  $K_n = 0,2224$ . Samuti ka  $\chi^2$ -testi korral on kõige parem tulemus log-normaalset jaotusel ning kõige halvem Waldi jaotusel. Antud andmemahu korral on K-S testi kriitiliseks väärtuseks 0,0041, millest kõik leitud teststatistikute väärtused on suuremad ning samuti ületavad  $\chi^2$ -teststatistikud kriitilisi väärtusi. Seega tuleb kummutada nullhüpotees ning võtta vastu hüpotees  $H_1$ , mis väidab jaotuste mittesobivust kahjuandmetele. Siiski on kahjuandmete esiosale jaotuste sobitamisel kõigi jaotuste teststatistikute väärtused paranenud, välja arvatud Pareto jaotus, mida ei olnud võimalik andmetele sobitada. Kui kogu kahjuandmestikule sobitamise korral andis kõige kehvema tulemuse gammajaotus, siis nüüd annab gammajaotus oluliselt parema sobivuse. Seda fakti seletab see, et gammajaotus ei ole raske sabaga jaotus ning kahjuandmete esiosa korral on raske saba eemaldatud.

Tabel 7: Sobivusnäitajate väärtused kahjuandmestku esiosal.

Jaotus	$K_n$	$K_n - \text{kriitiline}$	$\chi^2$	$\chi^2 - \text{kriitiline}$
<b>Gamma</b>	0,1018	0,0041	22554,36	160,9148
<b>Waldi</b>	0,2224	0,0041	35709,32	160,9148
<b>Log-normaalne</b>	0,0434	0,0041	10999,04	160,9148
<b>Pareto</b>	NaN	0,0041	NaN	160,9148
<b>Weibulli</b>	0,0841	0,0041	24413,36	160,9148

Tuginedes eelpool toodud analüüsile saame väita, et ükski toodud jaotustest ei kirjelda kahjuandmestiku esiosa hästi, kuid parima sobivuse annab siiski log-normaalne jaotus parameetritega  $\mu = 6,1634$  ning  $\sigma = 0,8732$ . Joonisel (17) on parema ülevaate saamiseks välja toodud kahjude esiosa empiiriline jaotus ning sobitatud log-normaalne jaotuse tihedusfunktsioon.



Joonis 17: Kaskokahjude empiiriline jaotus kahjuandmete esiosa korral ning sobitatud log-normaalne jaotus.

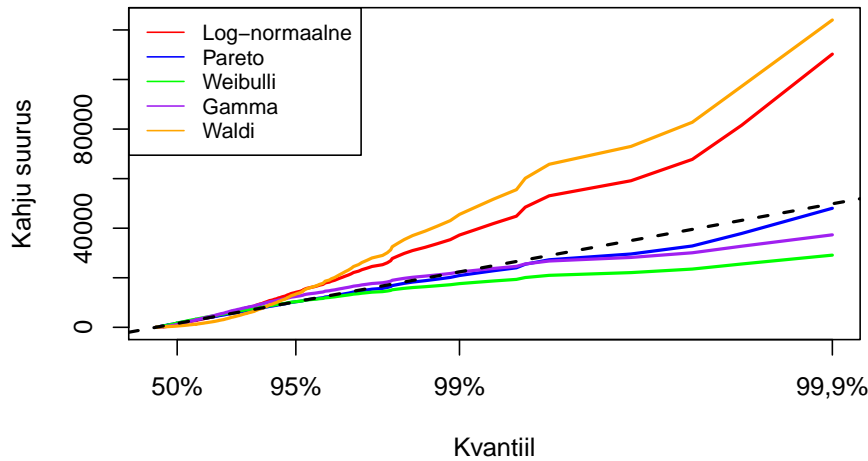
### 3.4.2 Jaotuste sobitamine kahjuandmestiku sabaosale

Järgnevalt sobitame jaotusi kahjuandmete sabaosale. Tabelis 8 on toodud sobitatud jaotuste suurima tõepära meetodil leitud parameetrite hinnangud.

Tabel 8: Parameetrite hinnangud kahjuandmestiku sabaosal.

Jaotus	Parameetrite hinnangud	
<b>Gamma</b>	$\alpha=0,4152$	$\lambda=0,0001$
<b>Waldi</b>	$\mu=3037,4920$	$\lambda=285,1119$
<b>Log-normaalne</b>	$\mu=7,2183$	$\sigma=1,4214$
<b>Pareto</b>	$\alpha=3,4206$	$\beta=7352,6880$
<b>Weibulli</b>	$\alpha=0,8074$	$\lambda=2659,2210$

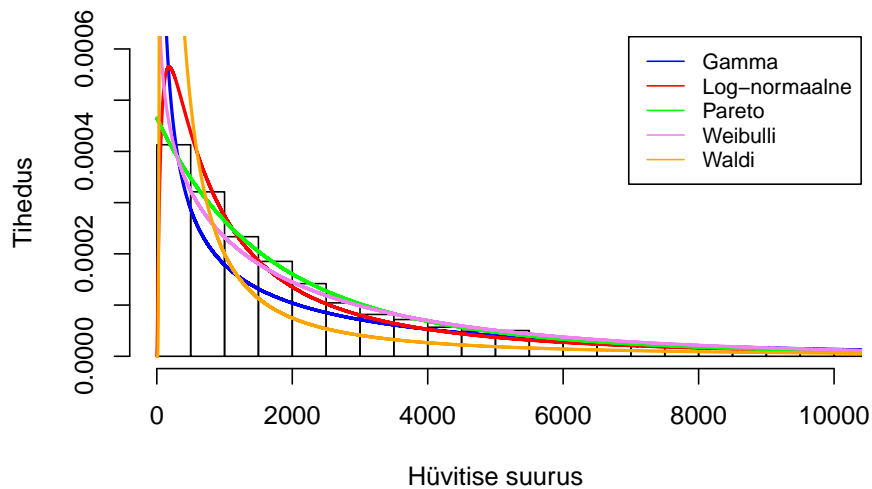
Joonisel (18) on kujutatud jaotuste kvantiilid võrreldes kahjuandmestiku sabaosa kvantiilidega. Jooniselt näeme, et kuni 95% kvantiilini järgivad peaaegu kõik jaotused kahjuandmestiku kvantiile. Alates 95% kvantiilist on näha, et kõige paremini järgib kahjuandmestiku kvantiile Pareto jaotus. Kuna jooniselt on väga raske välja lugeda jaotuste kvantiilide ühtivust kuni 95% kvantiilini, siis on lisades välja toodud ka kvantiilide tabel 16 kahjuandmestiku sabaosa kohta. Tabelist näeme, et Pareto jaotus järgib kahjuandmestiku kvantiile väga hästi üle kogu andmestiku sabaosa.



Joonis 18: Kvantiil-kvantiil graafikud kahjuandmestiku sabaosale.

Joonisel (19) on näha kahjude empiirilisele jaotusele sobitatud jaotuste tihedusfunktsioonid. Kahjuandmestiku sabaossa jäävad kahjud, mis on suuremad kui 3231, 43

eurot ning nad moodustavad 5% kõikide kahjude arvust, mis omakorda moodustavad 33% kogukahjust. Selleks, et kahjudele saaks sobitada jaotusi on kahju suurused nihutatud nullpunkti. Visuaalsel hinnangul annab kõige parema sobivuse Pareto jaotus. Samuti annab hea sobivuse log-normaalne jaotus, mis ennustab väiksemaid kahjusid rohkem, kui hetkel andmestikus on. Lisaks saame jooniselt näha, et andmete sabaosale ei sobitu üldse Waldi ja gammajaotus.



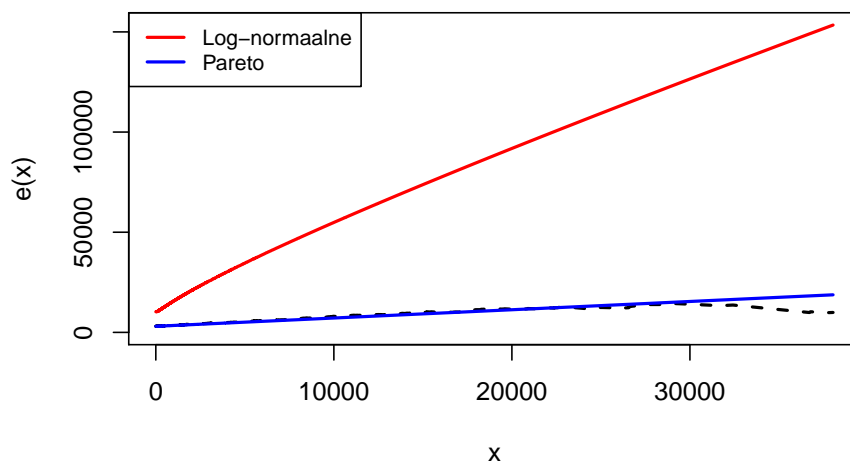
Joonis 19: Kaskokahjude empiiriline jaotus kahjuandmete sabaosa korral ning sobitatud jaotuste tihedusfunktsioonid.

Tabel 9 kinnitab fakti, et sabaosal annab kõige parema sobivuse Pareto jaotus, mille K-S teststatistiku väärtuseks on  $K_n = 0,0236$ . Antud andmemahu korral on K-S testi kriitiliseks väärtuseks 0,0178 ehk selle põhjal tuleks vastu võtta hüpotees  $H_1$ , mis väidab jaotuse mitte sobimist andmetele, aga siiski on Pareto jaotuse teststatistiku tulemus kriitilisele väärtusele väga lähedane. Lisaks on tabelis toodud  $\chi^2$ -testi tulemused. Pareto jaotuse korral on teststatistiku tulemus 249,1453 ning kriitiline väärtus 233,9943 ehk taaskord tuleb vastu võtta hüpotees  $H_1$ , kuigi tulemused on väga lähedased.

Tabel 9: Sobivusnäitajate väärtused kahjuandmestku sabaosale.

Jaotus	$K_n$	$K_n - \text{kriitiline}$	$\chi^2$	$\chi^2 - \text{kriitiline}$
<b>Gamma</b>	0,1672	0,0178	1462,00	233,9943
<b>Waldi</b>	0,2975	0,0178	3584,051	233,9943
<b>Log-normaalne</b>	0,0523	0,0178	647,4359	233,9943
<b>Pareto</b>	0,0236	0,0178	249,1453	233,9943
<b>Weibulli</b>	0,0384	0,0178	380,7009	233,9943

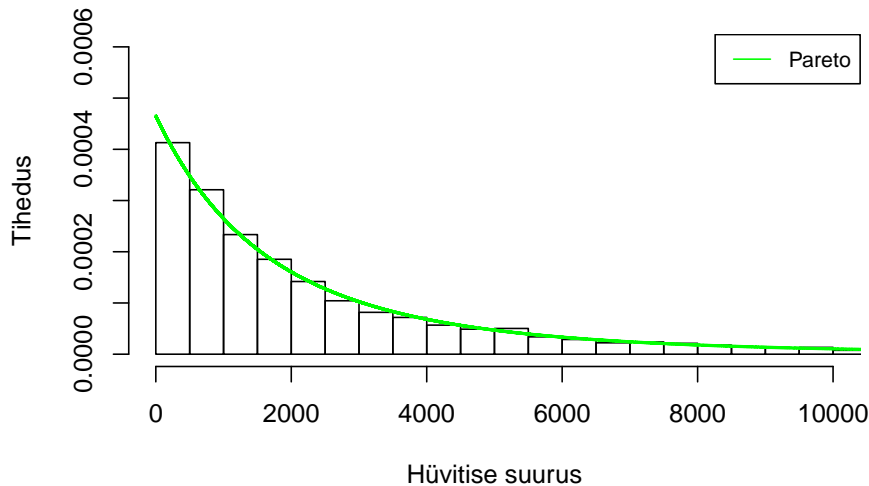
Jääkkeskmise meetodi rakendamisel saadud graafikud on toodud joonisel (20). Jooniselt näeme, et Pareto jaotuse jääkkeskmise graafik ühtib peaaegu kogu ulatuses kahjuandmestiku empiirilise jääkkeskmisega, mis annab kinnitust, et kahjuandmete sabaosale sobitub kõige paremini Pareto jaotus.



Joonis 20: Jääkkeskmise meetodi graafikud kahjuandmestiku sabaosa korral.

Tuginedes eelpool toodud analüüsile ei saa väita, et sabaosa kahjuandmed oleks Pareto jaotusest, aga Pareto jaotus parameetritega  $\alpha = 3,4206$  ning  $\beta = 7342,6880$  kirjeldab kahjusid andmestiku sabaosal hästi. Parema ülevaate saamiseks on joonisel (21) välja toodud kahjude empiiriline jaotus sabaosal ning sobitatud Pareto

jaotuse tihedusfunktsioon.



Joonis 21: Kaskokahjude empiiriline jaotus kahjuandmete sabaosa korral ning sobitatud Pareto jaotus.

### 3.5 Andmestiku jagamine kaheks kahjuliigiti

Järgmisena jagame kahjuandmestiku liigiti kaheks. Esimesse gruppi võtame ainult esiklaasikahjud ning teise kõik ülejäänud kahjud. Seda teeme seetõttu, et esiklaasikahjud käituvad üldjuhul teistmoodi, kui kõik ülejäänud kahjud. Esiklaasikahjudele ei rakendata reeglina omavastutust. Kahjusuurus, mis laekub kindlustusettevõttele esiklaasikahju korral, on suurus, mis tuleb kindlustusettevõttele tasuda. Esiklaasikahjud moodustavad 39,1% kõikidest kahjudest ehk nende osakaal andmestikus on väga suur. Lisaks on kahjuandmestiku esiosa empiirilise jaotuse korral näha väikeste kahjude korral suurt hüpet. Täpsemal uurimisel selgus, et tegemist on esiklaasikahjudega, mille suurus on 60 eurot ning kuna andmestikus oleva kahju kirjelduse põhjal ei ole võimalik midagi täpsemat teada saada, siis uuris autor selle kohta eksperdilt, kes töötas If kindlustuse sõidukikahjude grupijuhina. Selle tulemusena sai autor teada, et esiklaasikahjud jagunevad omakorda kaheks - esiklaasi vahetuseks ning esiklaasi takke paranduseks. Need 60 eurot maksvad kahjud ongi

täkke parandused, mille maksumus on fikseeritud töö tegija poolt.

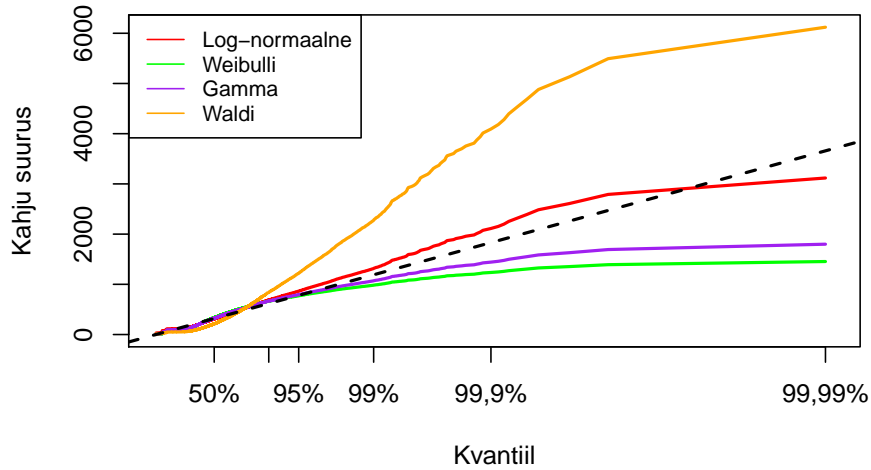
### 3.5.1 Jaotuste sobitamine esiklaasikahjudele

Järgnevalt sobitame jaotusi esiklaasikahjudele. Tabelis 10 on toodud sobitatud jaotuste suurima tõepära meetodil leitud parameetrite hinnangud.

Tabel 10: Parameetrite hinnangud esiklaasikahjude korral.

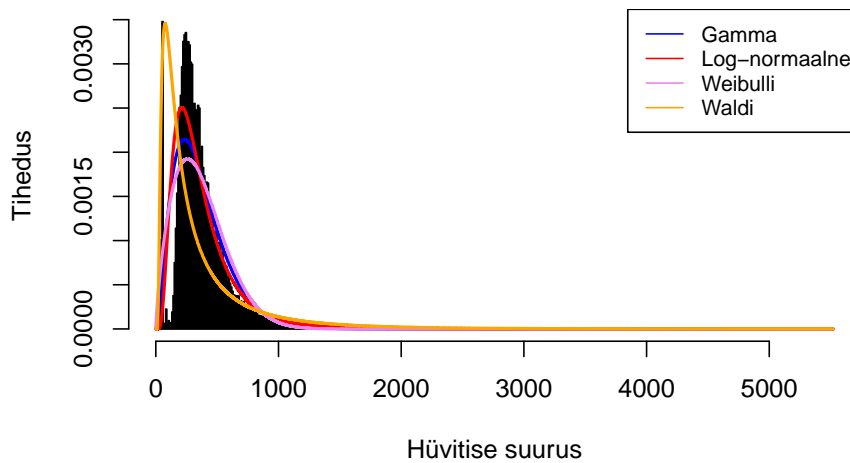
Jaotus	Parameetrite hinnangud	
<b>Gamma</b>	$\alpha=2,7406$	$\lambda=0,0074$
<b>Waldi</b>	$\mu=368,3056$	$\lambda=239,2591$
<b>Log-normaalne</b>	$\mu=5,7416$	$\sigma=0,6193$
<b>Pareto</b>	$\alpha=-1,1490$	$\beta=-791,4854$
<b>Weibulli</b>	$\alpha=1,7701$	$\lambda=414,8017$

Joonisel (22) on kujutatud jaotuste kvantiilid võrreldes kahjuandmestikus esinevate esiklaasikahjude kvantiilidega. Jooniselt näeme, et kuni 99% kvantiilini järgivad peaaegu kõik jaotused, välja arvatud Waldi jaotus, kahjuandmestiku kvantiile. Alates 99% kvantiilist on näha, et kõige paremini järgib kahjuandmestiku kvantiile log-normaalne jaotus. Kuna jooniselt on väga raske välja lugeda jaotuste kvantiilide ühtivust kuni 99% kvantiilini, siis on lisades välja toodud ka kvantiilide tabel 17 kahjuandmestikus esinevate esiklaasikahjude kohta. Tabelist näeme, et väiksemate kvantiilide korral on kooskõla parem Weibulli ja gammajaotuse korral ning suuremate kvantiilide puhul log-normaalne jaotuse korral.



Joonis 22: Kvantiil-kvantiil graafikud esiklaasikahjudele.

Joonisel (23) on näha esiklaasikahjude empiirilisele jaotusele sobitatud jaotuste tihedusfunktsioonid. Visuaalsel vaatlusel on raske öelda, milline jaotus võiks olla parim, sest suhteliselt sarnase sobivuse annavad log-normaalne, Weibulli ja gammajaotus. Kõige halvema sobivuse annab visuaalsel hinnangul Waldi jaotus, mis tabab ära küll väljalöögi empiirilises jaotuses, kuid ülejäänud jaotusele ei sobitu.



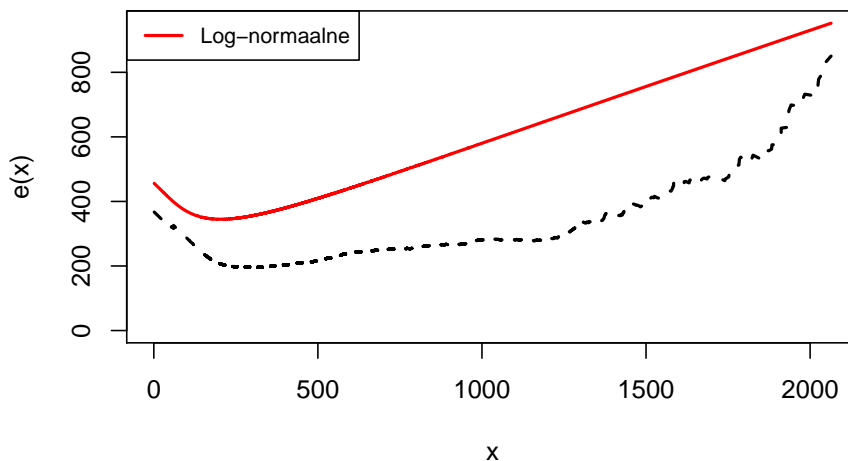
Joonis 23: Esiklaasikahjude empiiriline jaotus ning sobitatud jaotuste tihedusfunktsioonid.

Numbrilised näitajad jaotuste sobivuse kohta on toodud tabelis 11. Tabelist näeme, et K-S testi põhjal annab parima sobivuse log-normaalne jaotus, mille teststatistiku väärtus on  $K_n = 0,0961$ . Siinkohal tasub välja tuua, et log-normaalsega väga lähedase tulemuse andsid nii Weibulli kui ka gammajaotus.  $\chi^2$ -testi korral on vähim väärtus gammajaotusel, millele järgnevad Weibulli ja log-normaalne jaotus. Mõlema testi korral on teststatistikute väärtused mitu korda suuremad, kui kriitilised väärtused ning tuleb vastu võtta hüpotees  $H_1$ , mis väidab jaotuste mittedobivust kahjuandmetele.

Tabel 11: Sobivusnäitajate väärtused esiklaasikahjude korral.

Jaotus	$K_n$	$K_n - \text{kriitiline}$	$\chi^2$	$\chi^2 - \text{kriitiline}$
<b>Gamma</b>	0,103	0,00635	27661,91	522,5951
<b>Waldi</b>	0,3399	0,00635	57072,12	522,5951
<b>Log-normaalne</b>	0,0961	0,00635	36896,27	522,5951
<b>Pareto</b>	NaN	0,00635	NaN	522,5951
<b>Weibulli</b>	0,1062	0,00635	30734,57	522,5951

Jääkkeskmise meetodi rakendamisel saadud graafikud on toodud joonisel (24). Jooniselt näeme, et log-normaalse jaotuse jääkkeskmise graafik hindab üle kahjuandmete empiirilist jääkkeskmist, kuid oma kujult on graafikud siiski üsna sarnased.



Joonis 24: Jääkeskmise meetodi graafik esiklaasikahjude korral.

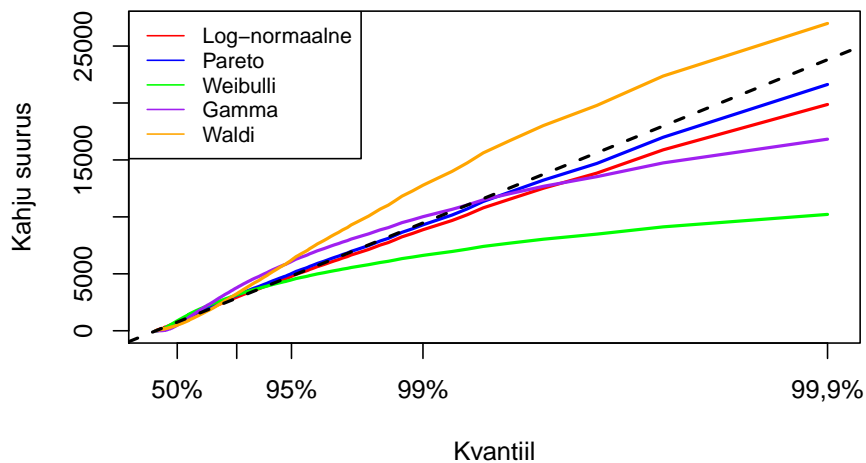
### 3.5.2 Jaotuste sobitamine ülejäänud kahjudele

Viimasena vaatleme kahjusid ilma esiklaasikahjudeta. Tabelis 12 on toodud sobitatud jaotuste suurima tõepära meetodil leitud parameetrite hinnangud.

Tabel 12: Parameetrite hinnangud ülejäänud kahjude korral.

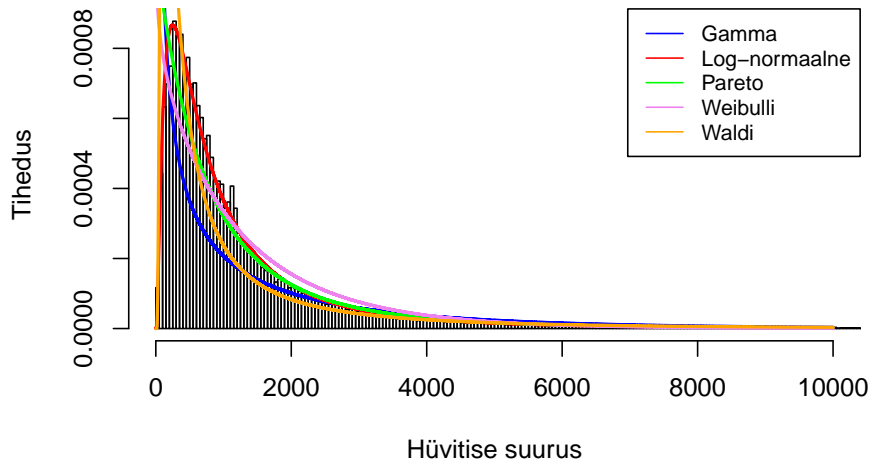
Jaotus	Parameetrite hinnangud	
<b>Gamma</b>	$\alpha=0,4021$	$\lambda=0,0003$
<b>Waldi</b>	$\mu=1339,3188$	$\lambda=362,1177$
<b>Log-normaalne</b>	$\mu=6,6354$	$\sigma=1,0555$
<b>Pareto</b>	$\alpha=3,3450$	$\beta=3140,7337$
<b>Weibulli</b>	$\alpha=0,9330$	$\lambda=1287,5481$

Joonisel (25) on kujutatud jaotuste kvantiilid võrreldes ülejäänud kahjude kvantiilidega. Jooniselt näeme, et nii log-normaalne kui ka Pareto jaotus järgivad üsnagi hästi kahjuandmete kvantiile. Kuna jooniselt on väga raske välja lugeda täpset jaotuste kvantiilide ühtivust kuni 99% kvantiilini, siis on lisades välja toodud ka kvantiilide tabel 18 ülejäänud kahjude kohta. Tabelist näeme, kuni 99% kvantiilini annab parema sobivuse log-normaalne jaotus ning peale seda Pareto jaotus.



Joonis 25: Kvantiil-kvantiil graafikud ülejäänud kahjudele.

Joonisel (26) näeme kahjude empiirilisele jaotusele sobitatud jaotuste tihedusfunktsioonid. Visuaalsel hinnangul annab parima sobivuse log-normaalne jaotus, mis tabab täpselt kahjude empiirilise jaotuse kuju. Kõige halvema sobivuse visuaalsel hinnangul annab gammajaotus, mis ei taba empiirilise jaotuse kuju väiksemate kahjude korral. Lisaks näeme, et varem väikeste kahjude korral esinenud väljalöök empiirilises jaotuses enam ei esine.



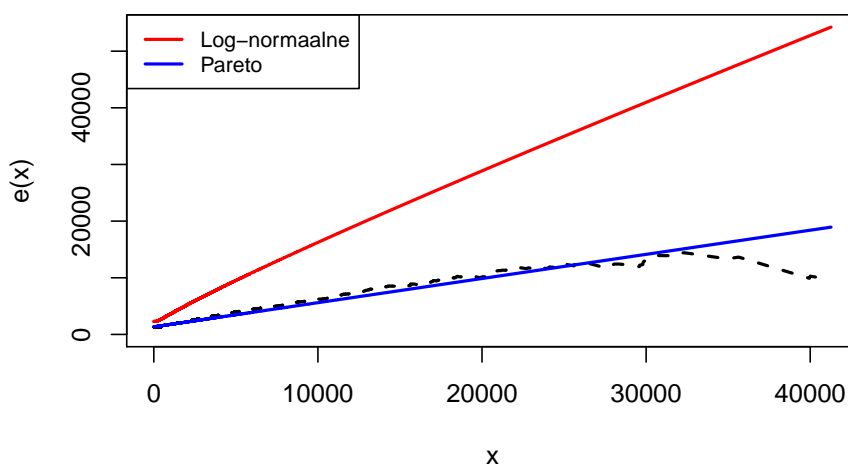
Joonis 26: Ülejäänud kahjude empiiriline jaotus ning sobitatud jaotuste tihe-  
dusfunktsioonid.

Visuaalseid tulemusi kinnitab ka tabel 13. K-S testi korral annab parima tulemuse log-normaalne jaotus, mille teststatistiku väärtus on  $K_n = 0,0171$ . Antud andme-  
mahu korral on K-S testi kriitiliseks väärtuseks  $0,00509$ , mis on väiksem kõigist  
teststatistikute väärtustest, kuid üsna lähedane log-normaalse testi tulemusega.  
 $\chi^2$ -testi korral annab samuti parima tulemuse log-normaalne jaotus, mille teststa-  
tistik on kõige väiksem, aga siiski ligi neli korda suurem kui kriitiline väärtus. Seega  
taaskord tuleb vastu võtta hüpotees  $H_1$ , mis väidab jaotuste mitesobivust andme-  
tele. Siiski tasub ära mainida, et log-normaalse jaotuse tulemus K-S testi korral on  
kriitilise väärtuse lähedal.

Tabel 13: Sobivusnäitajate väärtused ülejäänud kahjude korral.

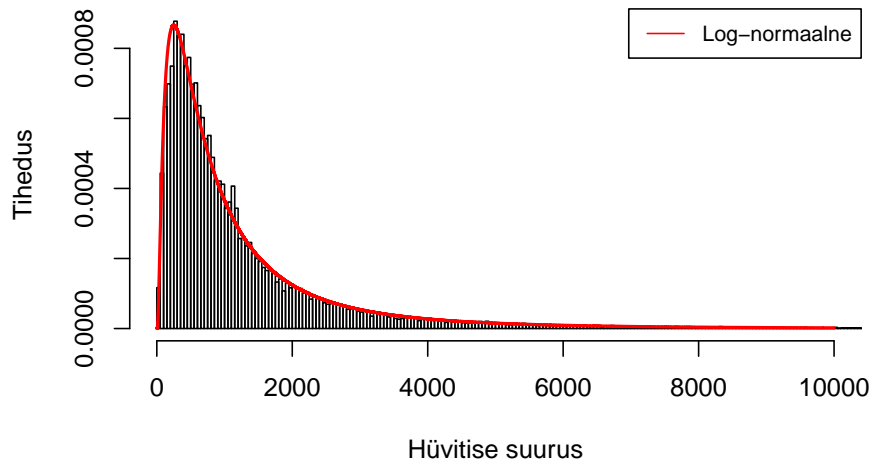
Jaotus	$K_n$	$K_n - \text{kriitiline}$	$\chi^2$	$\chi^2 - \text{kriitiline}$
<b>Gamma</b>	0,2622	0,00509	41555,88	1398,411
<b>Waldi</b>	0,1801	0,00509	20021,77	1398,411
<b>Log-normaalne</b>	0,0171	0,00509	5755,744	1398,411
<b>Pareto</b>	0,0939	0,00509	11729,04	1398,411
<b>Weibulli</b>	0,0751	0,00509	18250,94	1398,411

Jääkeskmise meetodi rakendamisel saadud graafikud on toodud joonisel (27). Joonis on väga sarnane kogukahjude korral saadud joonisega, mis on oodatav. Kuigi graafiku põhjal ning lisades toodud kvantiilide tabeli põhjal võiks sabaosal parema sobivuse anda Pareto jaotus, siis katsetades andmestikku jagada kaheks eespool kirjeldatud meetodiga ei andnud see paremaid tulemusi.



Joonis 27: Jääkeskmise meetodi graafik muude kahjude korral.

Tuginedes eespool toodud analüüsidele saame väita, et ükski toodud jaotustest ei kirjelda ülejäänud kahjusid piisavalt hästi, kuid parima sobivuse annab siiski log-normaalne jaotus parameetritega  $\mu = 6,5354$  ning  $\sigma = 1,0555$ . Joonisel (28) on parema ülevaate saamiseks välja toodud vaadeldavate kahjude empiiriline jaotus ning sobitatud log-normaalse jaotuse tihedusfunktsioon.



Joonis 28: Kaskokahjude empiiriline jaotus ülejäänud kahjude korral ning sobitatud log-normaalne jaotus.

## 4 Kokkuvõte

Kaskokindlustuse kahjuanalüüs näitas, et ühe kindla jaotusega ei õnnestunud leida kahjude kirjeldamiseks head mudelit. Parema tulemuse andis kahjude kirjeldamine kahe jaotusega, kus empiirilise jaotuse sabaosal andis parima lähenduse Pareto jaotus ning kahjuandmete esiosal log-normaalne jaotus. Soovitavaks kahe jaotuse eralduspunktiks oli 95% kvantiil. Kui kõigist kahjudest eraldada esiklaasikahjud, oli log-normaalne jaotus parimaks lähendavaks jaotuseks, kusjuures sobivus ülejäänud andmetele oli parem, kui selle jaotuse rakendamisel kogu algsele kahjuandmestikule.

Töös kasutatud jääkkeskmise meetod andis kasulikke suuniseid kahjujaotuse kirjeldamiseks kahe jaotuse abil.

Läbiviidud analüüsi põhjal võib soovitada firmal kahjude prognoosimisel jagada andmed kaheks 95% kvantiili kohalt ja rakendada sabaosal Pareto jaotust ning ülejäänud kahjuandmete kirjeldamiseks log-normaalset jaotust.

## 5 Kasutatud kirjandus

- [1] Hogg, R. V., Klugman, S. A. (1984). *Loss Distributions*. New York: John Wiley & Sons, Inc.
- [2] Tse, Y.-K. (2009). *Nonlife Actuarial Models. Theory, Methods and Evaluation*. Cambridge: Cambridge University Press.
- [3] Forbes, C., Evans, M., Hastings, N., Peacock, B. (2011). *Statistical Distributions. Fourth Edition*. New Jersey: John Wiley & Sons, Inc.
- [4] Hall, W. J., Wellner, J. A. (1981). Mean Residual Life. In: Eds. M. Csörgö et al. *Statistics and Related Topics*. Amsterdam: North-Holland, 169-184.
- [5] Kollo, T. (2004). *Monte Carlo meetodid*. Tartu: Tartu Ülikooli Kirjastus.
- [6] Freedman, D., Diaconis, P. (1981). On the histogram as a density estimator:  $L_2$  theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 453-476.
- [7] Embrechts, P., Klüppelberg, C., Mikosch, T. (2008). *Modelling Extremal Events for Insurance and Finance. Eighth Edition*. Heidelberg: Springer.

## 6 Lisad

### 6.1 Kvantiilide tabelid

Tabel 14: Kvantiilide võrdlus kogu kahjuandmestiku korral.

Kvantiil	Andmestik	LNorm	Pareto	Weibull	Gamma	Waldi
1%	50,00	51,31	6,33	5,58	0,00	41,95
5%	126,91	102,07	32,54	33,81	0,14	69,18
10%	191,10	147,27	67,46	74,91	1,32	94,35
20%	251,88	229,57	145,84	171,70	12,28	144,29
90%	2074,38	1956,25	2209,44	2265,08	2816,52	2326,26
95%	3231,43	2822,59	3297,64	3029,70	4337,35	3798,71
99%	7659,52	5614,68	7093,67	4873,00	8272,43	8512,48
99,9%	19936,49	12136,96	17890,80	7628,17	14401,68	17442,59

Tabel 15: Kvantiilide võrdlus kahjuandmestiku esiosa korral.

Kvantiil	Andmestik	LNorm	Pareto	Weibull	Gamma	Waldi
1%	46,57	62,30	NaN	16,61	13,57	38,96
5%	120,80	112,97	NaN	63,47	53,25	63,39
10%	186,93	155,14	NaN	114,71	98,05	85,51
20%	246,40	227,81	NaN	212,63	186,63	128,30
90%	1554,36	1454,58	NaN	1449,26	1494,65	1626,19
95%	2112,40	1997,62	NaN	1799,40	1907,55	2518,92
99%	2908,37	3622,06	NaN	2562,65	2853,64	5214,30
99,9%	3200,58	7057,39	NaN	3576,78	4190,91	10101,98

Tabel 16: Kvantiilide võrdlus kahjuandmestiku sabaosa korral.

Kvantiil	Andmestik	LNorm	Pareto	Weibull	Gamma	Waldi
1%	22,46	49,98	21,64	8,92	0,08	41,92
5%	103,93	131,68	111,09	67,15	4,03	71,32
10%	213,81	220,68	230,00	163,77	21,43	99,89
20%	478,71	412,41	495,64	414,87	114,76	160,33
90%	7148,83	8432,31	7061,45	7471,29	8522,19	6306,12
95%	10316,67	14132,18	10299,26	10350,28	12457,86	13408,86
99%	22375,01	37229,62	20904,65	17629,98	22316,31	45562,20
99,9%	49849,03	110262,16	48042,74	29131,22	37320,75	124025,03

Tabel 17: Kvantiilide võrdlus esiklaasikahjude korral.

Kvantiil	Andmestik	LNorm	Pareto	Weibull	Gamma	Waldi
1%	38,70	73,76	NaN	30,85	47,07	30,83
5%	60,00	112,50	NaN	77,46	92,41	48,92
10%	180,73	140,88	NaN	116,33	127,07	64,69
20%	220,97	185,00	NaN	177,76	180,69	93,92
90%	617,01	689,03	NaN	664,46	666,51	842,75
95%	780,63	862,89	NaN	770,97	793,50	1219,29
99%	1190,01	1315,99	NaN	982,96	1070,11	2277,35
99,9%	1830,84	2112,08	NaN	1236,00	1442,06	4092,61
99,99%	3658,52	3117,68	NaN	1454,12	1799,04	6120,71

Tabel 18: Kvantiilide võrdlus ülejäänud kahjude korral.

Kvantiil	Andmestik	LNorm	Pareto	Weibull	Gamma	Waldi
1%	62,50	65,36	9,45	9,30	0,03	50,93
5%	134,41	134,19	48,53	53,36	1,44	84,47
10%	208,10	196,91	100,50	115,41	8,07	115,76
20%	328,10	313,27	216,66	257,97	45,62	178,53
90%	2867,18	2945,45	3110,74	3147,68	3779,55	3249,42
95%	4323,99	4322,00	4550,15	4173,35	5554,72	5461,02
99%	9467,82	8873,04	9302,53	6616,65	10016,32	12779,54
99,9%	23804,52	19871,33	21626,98	10218,21	16823,70	26987,10

## 6.2 Kolmogorov-Smirnovi teststatistikute varieeruvus

Tabel 19: Kolmogorov-Smirnovi teststatistikute varieeruvus kogu kahjuandmestiku korral.

Andmestik	Jaotus	Keskmine	Min	Max
Kogu andmestik	Log-normaalne	0,0547	0,0518	0,0567
	Pareto	0,1525	0,1497	0,1566
	Weibulli	0,1224	0,1189	0,1259
	Gamma	0,3721	0,3539	0,3883
	Waldi	0,1833	0,1769	0,1893

Tabel 20: Kolmogorov-Smirnovi teststatistikute varieeruvus kahjuandmestiku esiosa korral.

Andmestik	Jaotus	Keskmine	Min	Max
Esiosa	Log-normaalne	0,0434	0,0397	0,0468
	Pareto	NaN	NaN	NaN
	Weibulli	0,0841	0,0791	0,0903
	Gamma	0,1018	0,0988	0,1048
	Waldi	0,2224	0,2131	0,2268

Tabel 21: Kolmogorov-Smirnovi teststatistikute varieeruvus kahjuandmestiku sabaosa korral.

Andmestik	Jaotus	Keskmine	Min	Max
Sabaosa	Log-normaalne	0,0523	0,0337	0,0807
	Pareto	0,0236	0,0089	0,0436
	Weibulli	0,0384	0,0248	0,0591
	Gamma	0,1672	0,1302	0,2078
	Waldi	0,2975	0,2619	0,3343

Tabel 22: Kolmogorov-Smirnovi teststatistikute varieeruvus esiklaasikahjude korral.

Andmestik	Jaotus	Keskmine	Min	Max
Esiklaas	Log-normaalne	0,0961	0,0905	0,1065
	Pareto	NaN	NaN	NaN
	Weibulli	0,1062	0,0993	0,1133
	Gamma	0,103	0,0936	0,1118
	Waldi	0,3399	0,3349	0,3496

Tabel 23: Kolmogorov-Smirnovi teststatistikute varieeruvus ülejäänud kahjude korral.

Andmestik	Jaotus	Keskmine	Min	Max
Ülejäänud	Log-normaalne	0,0171	0,0143	0,0251
	Pareto	0,0939	0,0887	0,1008
	Weibulli	0,0751	0,0696	0,0839
	Gamma	0,2622	0,2394	0,292
	Waldi	0,1801	0,1677	0,1926

Tabel 24: Kolmogorov-Smirnovi teststatistikute varieeruvus 95% kvantiililt jagamise korral.

<b>Kvantiil</b>	<b>Jaotus</b>	<b>Liik</b>	<b>Keskmine</b>	<b>Min</b>	<b>Max</b>
<b>95%</b>	<b>Log-normaalne</b>	Esiosa	0,0436	0,0387	0,0477
		Sabaosa	0,0507	0,0302	0,0765
	<b>Pareto</b>	Esiosa	NaN	NaN	NaN
		Sabaosa	0,0258	0,0102	0,0556
	<b>Weibulli</b>	Esiosa	0,0846	0,0774	0,0885
		Sabaosa	0,0406	0,0242	0,0660
	<b>Gamma</b>	Esiosa	0,1019	0,0980	0,1064
		Sabaosa	0,1657	0,1231	0,2131
	<b>Wald</b>	Esiosa	0,2219	0,2161	0,2301
		Sabaosa	0,2966	0,2398	0,3361

Tabel 25: Kolmogorov-Smirnovi teststatistikute varieeruvus 96% kvantiililt jagamise korral.

<b>Kvantiil</b>	<b>Jaotus</b>	<b>Liik</b>	<b>Keskmine</b>	<b>Min</b>	<b>Max</b>
<b>96%</b>	<b>Log-normaalne</b>	Esiosa	0,0445	0,0394	0,0496
		Sabaosa	0,0543	0,0285	0,0743
	<b>Pareto</b>	Esiosa	NaN	NaN	NaN
		Sabaosa	0,0260	0,0109	0,0510
	<b>Weibulli</b>	Esiosa	0,0869	0,0823	0,0912
		Sabaosa	0,0357	0,0227	0,0560
	<b>Gamma</b>	Esiosa	0,1153	0,1115	0,1195
		Sabaosa	0,1586	0,0913	0,1941
	<b>Wald</b>	Esiosa	0,2169	0,2099	0,2247
		Sabaosa	0,3583	0,2188	0,4446

Tabel 26: Kolmogorov-Smirnovi teststatistikute varieeruvus 97% kvantiililt jagamise korral.

<b>Kvantiil</b>	<b>Jaotus</b>	<b>Liik</b>	<b>Keskmine</b>	<b>Min</b>	<b>Max</b>
<b>97%</b>	<b>Log-normaalne</b>	Esiosa	0,0457	0,0424	0,0491
		Sabaosa	0,0502	0,0259	0,0842
	<b>Pareto</b>	Esiosa	0,1289	0,1260	0,1315
		Sabaosa	0,0306	0,0137	0,0523
	<b>Weibulli</b>	Esiosa	0,0901	0,0863	0,0971
		Sabaosa	0,0392	0,0218	0,0547
	<b>Gamma</b>	Esiosa	0,1326	0,1282	0,1354
		Sabaosa	0,1534	0,1072	0,1998
	<b>Wald</b>	Esiosa	0,2106	0,2045	0,2181
		Sabaosa	0,3291	0,2445	0,3928

Tabel 27: Kolmogorov-Smirnovi teststatistikute varieeruvus 98% kvantiililt jagamise korral.

<b>Kvantiil</b>	<b>Jaotus</b>	<b>Liik</b>	<b>Keskmine</b>	<b>Min</b>	<b>Max</b>
<b>98%</b>	<b>Log-normaalne</b>	Esiosa	0,0468	0,0434	0,0520
		Sabaosa	0,0676	0,0395	0,1068
	<b>Pareto</b>	Esiosa	0,1307	0,1286	0,1339
		Sabaosa	0,0296	0,0116	0,0663
	<b>Weibulli</b>	Esiosa	0,0942	0,0916	0,0985
		Sabaosa	0,0429	0,0217	0,0812
	<b>Gamma</b>	Esiosa	0,1556	0,1522	0,1594
		Sabaosa	0,1374	0,0777	0,1845
	<b>Wald</b>	Esiosa	0,2036	0,1982	0,2098
		Sabaosa	0,3384	0,2703	0,3921

Tabel 28: Kolmogorov-Smirnovi teststatistikute varieeruvus 99% kvantiililt jagamise korral.

<b>Kvantiil</b>	<b>Jaotus</b>	<b>Liik</b>	<b>Keskmine</b>	<b>Min</b>	<b>Max</b>
<b>99%</b>	<b>Log-normaalne</b>	Esiosa	0,0493	0,0477	0,0528
		Sabaosa	0,0671	0,0298	0,1160
	<b>Pareto</b>	Esiosa	0,1338	0,1308	0,1370
		Sabaosa	0,0443	0,0192	0,0912
	<b>Weibulli</b>	Esiosa	0,1005	0,0988	0,1039
		Sabaosa	0,0517	0,0232	0,0991
	<b>Gamma</b>	Esiosa	0,1936	0,1907	0,1976
		Sabaosa	0,1379	0,0865	0,1877
	<b>Wald</b>	Esiosa	0,1962	0,1873	0,2067
		Sabaosa	0,3369	0,2312	0,4405

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Hardo Niit,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose “Kasko-kindlustuse kahjude modelleerimine”, mille juhendaja on Tõnu Kollo, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Hardo Niit

25.05.2021