

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Anet Ilustrumm
**Raske sabaga Lambert'i W teisendus ja selle
kasutamine kahjujaotuste hindamisel**

Matemaatilise statistika eriala

Bakalaureusetöö (9 EAP)

Juhendajad: Meelis Käärik, PhD
Anne Selart, MSc

TARTU 2025

RASKE SABAGA LAMBERT'I W TEISENDUS JA SELLE KASUTAMINE KAHJUJAOTUSTE HINDAMISEL

Bakalaureusetöö

Anet Ilustrumm

Lühikokkuvõte

Antud bakalaureusetöö eesmärk on tutvustada raske sabaga Lambert'i W teisendust ning uurida selle matemaatilist kirjeldust nii üldkujul kui ka juhul, kus algjaotuseks on eksponent- või normaaljaotus. Lisaks vaadatakse juhtu, kus juhuslik suurus Y on määratud eraldi vasaku ja parema saba parameetrite kaudu. Töö praktilises osas rakendatakse raske sabaga Lambert'i $W \times \text{Exp}(\lambda)$ ja kahe erineva raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ jaotusi neljale kahjuandmestikule ning uuritakse jaotuste sobivust. Saadud tulemusi võrreldakse artiklis Käärrik *et al.* (2023) saadud Lambert'i W jaotuste ja teiste tuntud 19 kahjujaotusega (Eling, 2012). Töö teoreetilise osa aluseks on võetud Goerg'i artikkel pealkirjaga „The Lambert Way to Gaussianize Heavy-Tailed Data with the Inverse of Tukey's h Transformation as a Special Case“ (Goerg, 2015b).

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

Märksõnad: juhuslikud suurused, kahjukindlustus, tõenäosusjaotused, asümmeetria

THE HEAVY-TAILED LAMBERT W TRANSFORMATION AND ITS APPLICATION IN MODELING LOSS DISTRIBUTIONS

Bachelor thesis

Anet Ilustrumm

Abstract

The aim of this bachelor's thesis is to introduce the heavy-tailed Lambert W transformation and to explore its mathematical formulation both in general and in the specific cases where the input distribution is exponential or normal. Furthermore, a case is considered where the random variable Y is defined using separate parameters for the left and right tails. In the practical part of the thesis, the heavy-tailed Lambert $W \times \text{Exp}(\lambda)$ and heavy-tailed Lambert $W \times N(\mu_X, \sigma_X)$ distributions are applied to four loss datasets, and the goodness-of-fit of these distributions is examined. The obtained results are compared with those of the Lambert W distributions and 19 other well-known loss distributions presented in Käärik *et al.* (2023) and Eling (2012). The thesis is based on the article by Goerg (2015b) titled „The Lambert Way to Gaussianize Heavy-Tailed Data with the Inverse of Tukey's h Transformation as a Special Case“.

CERCS research specialisation: P160 Statistics, operation research, programming, actuarial mathematics

Key Words: *random variables, non-life insurance, probability distributions, asymmetry*

Sisukord

Sissejuhatus	4
1 Lambert'i W funktsioon	6
1.1 Tähistused	6
1.2 Abitulemused	9
2 Raske sabaga Lambert'i W juhuslikud suurused	12
2.1 Definiitsioonid	12
2.2 Raske sabaga Lambert'i W eksponentjaotus	16
2.3 Raske sabaga Lambert'i W normaaljaotus	19
2.4 Asümmeetriline kahe sabaga juht	21
3 Jaotuste sobitamine andmetele	24
3.1 Parameetrite hindamine	24
3.2 Jaotuste sobitamine kahjuandmetele	25
Kokkuvõte	30
Kasutatud allikad	32
Lisa A Sobitatud jaotuste AIC sobivusnäitajad	34
A.1 USA andmetikele sobitatud 21 jaotuse AIC sobivusnäitajad	34
A.2 Taani andmetikele sobitatud 21 jaotuse AIC sobivusnäitajad	35

Sissejuhatus

Kindlustus- ja finantsmatemaatikas tuleb tihti ette olukordi, kus andmed on tugeva asümmeetriaga või raske sabaga jaotusest. Sellised andmed erinevad oluliselt klassikalisest normaaljaotusest, mille eeldused (sümmeetrilisus ja kerge saba) tihti ei kehti. See muudab andmete analüüsi ja modelleerimise keerulisemaks, kuna traditsiooniliste meetodite, näiteks tavalise normaaljaotuse, kasutamine võib anda eksitavaid tulemusi. Probleemi lahendamiseks on välja pakutud mitmesuguseid asümmeetrilisi ja raske sabaga jaotusi, nagu näiteks asümmeetriline t -jaotus ja log-normaalne jaotus. Üheks võimalikuks tööriistaks on ka selliste andmete modelleerimisel saanud Lambert'i W teisendus, mis võimaldab olemasoleva sümmeetrilise jaotuse muuta asümmeetriliseks ja/või raske sabaga jaotuseks.

Antud bakalaureusetöö on jätk T. Puhkimi magistritööle (Puhkim, 2016). Töö eesmärk on tutvustada raske sabaga Lambert'i W teisendust ning uurida selle matemaatilist kirjeldust nii üldkujul kui ka juhul, kus algjaotuseks on normaaljaotus või eksponentjaotus. Praktilises osas analüüsitakse, kui hästi saadud jaotused sobivad kahjuandmetele. Seejuures võrreldakse, kas raske sabaga Lambert'i W teisendusel põhinev jaotus sobiks praktiliseks rakendamiseks paremini kui Lambert'i W jaotused ja teised traditsioonilised kahjujaotused.

Bakalaureusetöö koosneb kolmest peatükist. Esimeses peatükis defineeritakse Lambert'i W funktsioon ning skaalapere ja asukoha-skaalapere mõisted. Samuti defineeritakse funktsioon W_δ , mida kasutatakse edaspidi lihtsuse huvides. Lisaks esitatakse mõned abitulemused funktsiooni W_δ omaduse ja diferentseeruvuse kohta ning näidatakse nende kehtivust. Neid tulemusi kasutatakse hiljem töös korduvalt. Peatükk on kirjutatud tuginedes allikatele Goerg (2015b), Goerg (2015a), Puhkim (2016) ning Casella ja Berger (2001).

Teises peatükis tutvustatakse Goerg'i artiklis välja toodud ideid juhuslike suuruste teisendamisest raske sabaga Lambert'i W juhuslikuks suuruseks (Goerg, 2015b). Positiivse δ korral tutvustatakse raske sabaga Lambert'i W juhusliku suuruse Y

jaotus- ja tihedusfunktsioone kolme jaotustüübi (mittetsentraalse ja mitteskaleeritud, skaleeritud ja asukoha-skaala) korral. Saadud valemeid rakendatakse eksponentjaotuse ja normaaljaotuse puhul. Lugemise lihtsustamiseks on esitatud graafikud erinevate sabaparameetritega jaotus- ja tihedusfunktsioonidest. Viimaks käsitletakse olukorda, kus juhuslik suurus Y on määratud eraldi vasaku ja parema saba parameetrite kaudu.

Kolmandas peatükis antakse lühike ülevaade raske sabaga Lambert'i $W \times F_X$ jaotuse parameetrite hindamise meetodist. Seejärel sobitatakse raske sabaga Lambert'i W jaotusi kahjuandmetele ja hinnatakse vastavad jaotuste parameetrid. Seejärel võrreldakse saadud jaotuste sobivusnäitajaid varasemalt saadud Lambert'i W jaotuste sobivusnäitajatega, et teada saada, kas raske sabaga mudelid pakuvad paremat sobivust reaalse andmestiku puhul. Viimaks võrreldakse saadud tulemusi ka teiste kahjujaotustega.

Bakalaureusetöö praktilise osa ja jooniste tegemiseks on kasutatud statistikatarkvara R. Raske sabaga Lambert'i W jaotus- ja tihedusfunktsioonide leidmiseks on kasutatud R paketti *LambertW* (Goerg, 2025) ning parameetrite hindamiseks kasutatakse paketti *fitdistrplus* (Delignette-Muller *et al.*, 2025).

1 Lambert'i W funktsioon

Selles peatükis defineerime Lambert'i W funktsiooni ja vaatame mõningaid raske sabaga Lambert'i W funktsiooni abitulemusi, mida meil läheb järgneva peatüki teisenduste aruteludes vaja, tuginedes Goerg'i artiklile (Goerg, 2015b) ja sealsele täiendava materjali dokumendile (Goerg, 2015a). Lisaks seletame lahti mõned töös vajalikud mõisted.

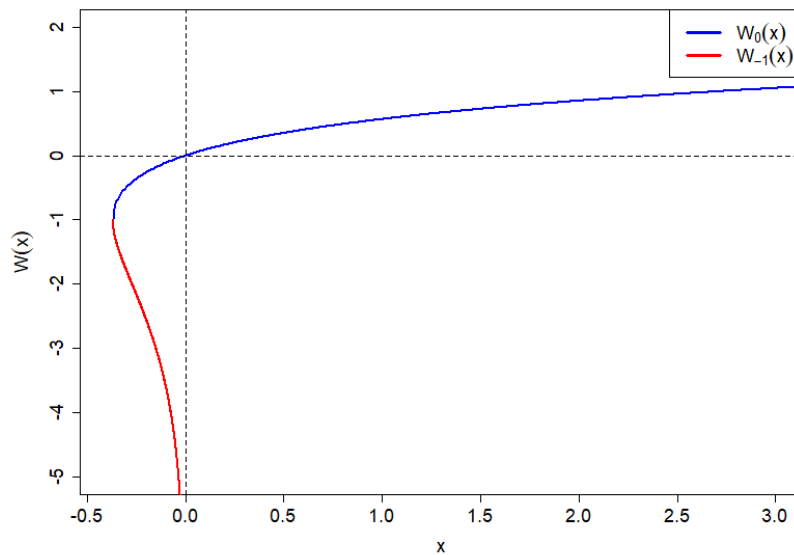
1.1 Tähistused

See bakalaureusetöö põhineb ühel olulisel definitsioonil.

Definitsioon 1. Lambert'i W funktsioon on defineeritud kui funktsioon, mis rahuldab võrrandit

$$W(x)e^{W(x)} = x, \quad x \in \left[-\frac{1}{e}, \infty\right). \quad (1)$$

Lambert'i W funktsioon pole üheselt määratud ning see muudab kogu järgneva töö selle funktsiooniga tunduvalt keerulisemaks (s.t funktsiooni peab vaatama arvutustes erinevate harude korral, vt joonis 1). Täpsemalt Lambert'i W funktsiooni ajaloo ning muu kohta saab lugeda Puhkimi magistritööst, kus on välja toodud ka mõned Lambert'i W funktsiooni omadused jms (Puhkim, 2016). Funktsiooni Lambert'i W mitteühesusest tulenevate probleemide vältimiseks võttis Goerg (2015b) kasutusele Lambert'i W funktsioonil põhineva aga erineva teisenduse, mida tähistame kui W_δ .



Joonis 1: Lambert'i W funktsioon.

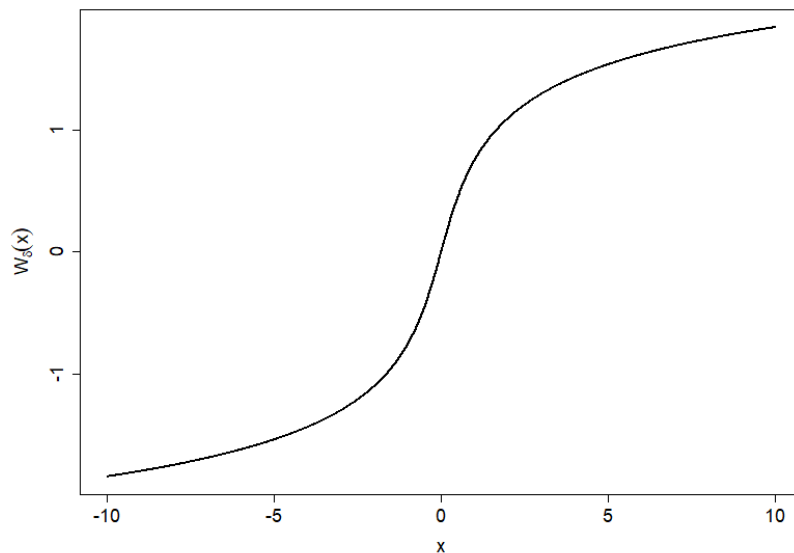
Edaspidi kasutame kogu bakalaureusetöö vältel lihtsuse huvides tähistust

$$W_\delta(z) = \operatorname{sgn}(z) \sqrt{\frac{W(\delta z^2)}{\delta}}, \quad z \in \mathbb{R},$$

kus funktsioon $\operatorname{sgn}(z)$ on defineeritud järgmiselt:

$$\operatorname{sgn}(z) = \begin{cases} 1, & \text{kui } z > 0, \\ 0, & \text{kui } z = 0, \\ -1, & \text{kui } z < 0. \end{cases}$$

Funktsioon W_δ on kogu reaalteljel rangelt kasvav ning bijektiivne (vt joonis 2). Just bijektiivsuse omadus muudab meie bakalaureusetöös edasised arvutused tunduvalt lihtsamaks.



Joonis 2: W_δ funktsioon.

Selleks, et järgmises peatükis olevaid definitsioone paremini mõista, seletame lahti kõigepealt mõned mõisted Casella ja Bergeri raamatu põhjal (Casella ja Berger, 2001, lk 116-121).

Definitsioon 2. Olgu $f(x)$ mistahes tihedusfunktsioon. Siis mistahes $\sigma > 0$ korral nimetatakse parameetriga σ indekseeritud tihedusfunktsioonide $\frac{1}{\sigma}f(\frac{x}{\sigma})$ peret skaalapereks, kus $f(x)$ on standardne tihedusfunktsioon ja σ on pere skaalaparameter.

See tähendab, et skaalaparameter $\sigma > 1$ määrab, kui „laiaks venitab“ ta $f(x)$ graafiku ja $\sigma < 1$ määrab vastupidi, kui „kitsaks tõmbab“ ta $f(x)$ graafiku. Hea näide sellest on skaalapere $N(0, \sigma)$ - kõigil nende jaotuste graafikutel on sarnane kuju, kuid hajuvus on erinev.

Definitsioon 3. Olgu $f(x)$ mistahes tihedusfunktsioon. Siis mistahes μ , $-\infty < \mu < \infty$ ja $\sigma > 0$ korral nimetatakse parameetritega (μ, σ) indekseeritud tihedusfunktsioonide $\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$ peret asukoha-skaalapereks, kus $f(x)$ on standardne tihedusfunktsioon, μ on asukohaparameter ja σ on skaalaparameter.

See tähendab, et asukohaparameter μ määrab, kuhu jaotus nihkub ning skaalaparameter määrab, kui „kitsas“ või „lai“ jaotus on. Hea näide sellest on normaaljaotus $N(\mu, \sigma)$.

1.2 Abitulemused

Esiteks näitame Lambert'i W funktsiooni omaduse kehtivust, mida kasutame hilisemas peatükis.

Lemma 1. Kehtib omadus

$$\frac{W_\delta(z)}{z} = e^{-\frac{\delta}{2}W_\delta^2(z)}. \quad (2)$$

Tõestus. Lamberti' W funktsiooni definitsiooni põhjal kehtib võrdus

$$\begin{aligned} W(\delta z^2)e^{W(\delta z^2)} &= \delta z^2 \Leftrightarrow \\ \frac{W(\delta z^2)}{\delta z^2}e^{W(\delta z^2)} &= 1 \Leftrightarrow \\ \frac{W(\delta z^2)}{\delta z^2} &= e^{-W(\delta z^2)}. \end{aligned}$$

Kui nüüd tõsta võrduse mõlemad pooled astmesse $\frac{1}{2}$, saame

$$\begin{aligned} \frac{W(\delta z^2)}{\delta z^2} &= e^{-W(\delta z^2)} \Leftrightarrow \\ \frac{\sqrt{W(\delta z^2)/\delta}}{|z|} &= e^{-\frac{W(\delta z^2)}{2}} = e^{-\frac{\delta}{2}\frac{W(\delta z^2)}{\delta}} = e^{-\frac{\delta}{2}W_\delta^2(z)}. \end{aligned}$$

Nüüd saame ära kasutada teadmist, et $|z| = \text{sgn}(z)z$, kui $z \neq 0$. Seega oleme saanud, et

$$\begin{aligned} \frac{\sqrt{W(\delta z^2)/\delta}}{\text{sgn}(z)z} &= e^{-\frac{\delta}{2}W_\delta^2(z)} \Leftrightarrow \\ \frac{W_\delta(z)}{z} &= e^{-\frac{\delta}{2}W_\delta^2(z)}. \end{aligned}$$

Kui $z = 0$, siis vaatame avaldist $\frac{\sqrt{W(\delta z^2)/\delta}}{z}$ eraldi piiril $z \rightarrow 0$. Kõigepealt asendame $x = \delta z^2$. Saame

$$\lim_{x \rightarrow 0^+} \sqrt{\frac{W(x)}{x}}.$$

Puhkim (2016) näitas, et Lambert'i W funktsiooni peaharu saab arendada nulli ümbruses Taylori ritta, niisiis saame

$$\lim_{x \rightarrow 0^+} \sqrt{\frac{W(x)}{x}} = \lim_{x \rightarrow 0^+} \sqrt{\frac{x - x^2 + 3x^3/2 - \dots}{x}} = \lim_{x \rightarrow 0^+} \sqrt{1 - x + \frac{3}{2}x^2 - \dots} = 1$$

ning teiselt poolt $e^{-\frac{\delta}{2}W_\delta^2(0)} = 1$, sest $W(0) = 1$. $\square$

Järgmiseks näitame, et funktsiooni W_δ saab diferentseerida. Funktsiooni W on võimalik diferentseerida, seega on võimalik diferentseerida ka funktsiooni W_δ . Meeldetuletuseks Puhkimi magistritööst funktsiooni W tuletis on

$$W'(x) = \frac{W(x)}{x(1 + W(x))}, \quad x \notin \left\{0, -\frac{1}{e}\right\}$$

(Puhkim, 2016).

Sõnastame funktsiooni W_δ diferentseeruvuse kohta lemma.

Lemma 2. Funktsiooni $W_\delta(z)$ tuletis z suhtes on

$$\frac{d}{dz} W_\delta(z) = \frac{W_\delta(z)}{z(1 + W(\delta z^2))}. \quad (3)$$

Tõestus. Enne kui saame näidata funktsiooni W_δ diferentseeruvust, on meil vaja näidata funktsiooni $\frac{W(\delta z^2)}{\delta}$ diferentseeruvust. Goerg (2015a) näitas oma täiendava materjali dokumendis olevas tõestuses, et

$$\frac{d}{dz} \frac{W(\delta z^2)}{\delta} = \frac{W'(\delta z^2)}{\delta} \cdot 2\delta z = \frac{W(\delta z^2)}{\delta z^2(1 + W(\delta z^2))} \cdot 2z = \frac{2W(\delta z^2)}{\delta z(1 + W(\delta z^2))}.$$

Seega nüüd saame diferentseerida funktsiooni $W_\delta(z)$.

$$\begin{aligned}
\frac{d}{dz} W_\delta(z) &= \frac{d}{dz} \left[\operatorname{sgn}(z) \sqrt{\frac{W(\delta z^2)}{\delta}} \right] \\
&= \operatorname{sgn}(z) \frac{1}{2} \left(\frac{W(\delta z^2)}{\delta} \right)^{-\frac{1}{2}} \cdot \frac{d}{dz} \frac{W(\delta z^2)}{\delta} \\
&= \operatorname{sgn}(z) \frac{1}{2} \left(\frac{W(\delta z^2)}{\delta} \right)^{-\frac{1}{2}} \cdot \frac{2W(\delta z^2)}{\delta z (1 + W(\delta z^2))} \\
&= \operatorname{sgn}(z) \left(\frac{W(\delta z^2)}{\delta} \right)^{\frac{1}{2}} \cdot \frac{1}{z(1 + W(\delta z^2))} \\
&= \frac{W_\delta(z)}{z(1 + W(\delta z^2))}.
\end{aligned}$$

□

Siinkohal tasub mainida, et Goerg'i täiendava materjali dokumendis on tehtud viga teisendustes (Goerg, [2015a](#)).

2 Raske sabaga Lambert'i W juhuslikud suurused

Järgnevas peatükis tutvustame raske sabaga Lambert'i W juhuslikke suursi ja tutvustame nende jaotus- ning tihedusfunktsiooni üldkuju, lisaks vaatame kahte konkreetset näidet eksponent- ja normaaljaotusega. Peatüki lõpus käsitletakse olukorda, kus juhuslik suurus Y on määratud eraldi vasaku ja parema saba parameetrite kaudu.

2.1 Definitsioonid

Selles alapeatükis toome välja Goerg'i artikli põhjal, kuidas defineeritakse raske sabaga Lambert'i W juhuslikke suursi erinevate jaotustüüpide korral (Goerg, 2015b).

Definitsioon 4. Kui U on pidev juhuslik suurus jaotusfunktsiooniga $F_U(u) = \mathbb{P}(U \leq u)$, $u \in \mathbb{R}$, tihedusfunktsiooniga $f_U(u)$, siis ütleme, et

$$Y := U \exp\left(\frac{\delta}{2}U^2\right), \quad \delta \in \mathbb{R}, \quad (4)$$

on mittetsentraalne, mitteskaleeritud raske sabaga Lambert'i $W \times F_U$ juhuslik suurus sabaparaameetriga δ .

Teame, et eksponentfunktsiooni väärtus on alati mittenegatiivne - järelikult U ja Y on sama märgiga. Kuna U^2 on alati mittenegatiivne, siis $\delta > 0$ korral $\exp\left(\frac{\delta}{2}U^2\right) > 1$ võimendab U väärtusi ja $\delta < 0$ korral $0 < \exp\left(\frac{\delta}{2}U^2\right) < 1$ vähendab U mõju. Kui $\delta = 0$, siis $\exp\left(\frac{\delta}{2}U^2\right) = 1$ ning järelikult teisendusest (4) saame samasuse $U \equiv Y$.

Kui sisendina kasutatakse skaalaperest pärit juhuslikku suurust X , siis tahaksime, et sama omadus oleks ka raske sabaga Lambert'i $W \times F_X$ juhuslikul suurusel. Sama põhimõtte kehtib ka asukoha-skaalaperest pärit oleva jaotusfunktsiooniga juhuslike

suuruste puhul. Küll aga teisendusel (4) selline omadus puudub. Seega esitame veel kaks definitsiooni.

Definitsioon 5. Kui X on pidev juhuslik suurus skaalaperest jaotusfunktsiooniga $F_X(x|\beta)$, kus β on F_X parameeter(vektor) ja σ_X on standardhälve, siis $U = \frac{X}{\sigma_X}$, seega $DU = \frac{1}{\sigma_X}DX = 1$, ning ütleme, et

$$Y := \left\{ U \exp\left(\frac{\delta}{2}U^2\right) \right\} \sigma_X = X \exp\left(\frac{\delta X^2}{2\sigma_X^2}\right), \quad \delta \in \mathbb{R}, \quad \sigma_X > 0 \quad (5)$$

on skaleeritud raske sabaga Lambert'i $W \times F_X$ juhuslik suurus parameetervektoriga $\theta = (\beta, \delta)$.

Definitsioon 6. Kui X on juhuslik suurus asukoha-skaalaperest jaotusfunktsiooniga $F_X(x|\beta)$, kus β on F_X parameeter(vektor), μ_X on keskväärus ja σ_X standardhälve, siis $U = \frac{X - \mu_X}{\sigma_X}$, seega $EU = \frac{EX - \mu_X}{\sigma_X} = 0$ ja $DU = \frac{DX}{\sigma_X} = 1$ ning ütleme, et

$$Y := \left\{ U \exp\left(\frac{\delta}{2}U^2\right) \right\} \sigma_X + \mu_X, \quad \delta \in \mathbb{R}, \quad \sigma_X > 0 \quad (6)$$

on asukoha-skaala raske sabaga Lambert'i $W \times F_X$ juhuslik suurus parameetervektoriga $\theta = (\beta, \delta)$.

Olgu X skaalaperest pärit mittenegatiivne juhuslik suurus jaotusfunktsiooniga F_X ja tihedusfunktsiooniga f_X ning olgu standardhälve σ_X . Kogu bakalaureusetöö vältel eeldame, et parameeter δ on mittenegatiivne. Raske sabaga Lambert'i W funktsioon on alati monotoonne ja üheselt määratud. Nüüd kasutame definitsioonis 5 defineeritud skaalaperest pärit juhuslikku suurust Y ning $U = \frac{X}{\sigma_X}$. Tähistame $z = \frac{y}{\sigma_X}$. Leiame raske sabaga Lambert'i $W \times F_X$ jaotusfunktsiooni F_Y jaotusfunktsiooni F_X kaudu järgmiselt

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \stackrel{(5)}{=} \mathbb{P}\left(U \exp\left(\frac{\delta}{2}U^2\right) \sigma_X \leq y \right) = \mathbb{P}\left(U \exp\left(\frac{\delta}{2}U^2\right) \leq z \right) \\ &= \mathbb{P}\left(\delta U^2 \exp(\delta U^2) \leq \delta z^2 \right) \stackrel{(1)}{=} \mathbb{P}(\delta U^2 \leq W(\delta z^2)) = \mathbb{P}\left(U^2 \leq \frac{W(\delta z^2)}{\delta} \right). \end{aligned}$$

Edasi vaatame, kui $z > 0$. Siinkohal me üldsust ei kitsenda, kuna X on eelduse põhjal mittenegatiivne, niisiis $z < 0$ korral on $F_Y(y) = 0$. Nüüd saame, et

$$F_Y(y) = \mathbb{P}\left(U^2 \leq \frac{W(\delta z^2)}{\delta}\right) = \mathbb{P}\left(U \leq \sqrt{\frac{W(\delta z^2)}{\delta}}\right) = F_X\left(\sqrt{\frac{W(\delta z^2)}{\delta}}\sigma_X\right).$$

Seega üldisemalt skaleeritud raske sabaga Lambert'i $W \times F_X$ juhusliku suuruse Y jaotusfunktsioon on

$$F_Y(y|\boldsymbol{\beta}, \delta) = F_X\left(\sqrt{\frac{W(\delta(y/\sigma_X)^2)}{\delta}}\sigma_X \middle| \boldsymbol{\beta}\right). \quad (7)$$

Juhusliku suuruse Y tihedusfunktsiooni leidmiseks võtame eelmisest avaldisest (7) tuletise y järgi, niisiis saame valemi (3) põhjal

$$\begin{aligned} f_X(y|\boldsymbol{\beta}, \delta) &= f_X\left(\sqrt{\frac{W(\delta(y/\sigma_X)^2)}{\delta}}\sigma_X \middle| \boldsymbol{\beta}\right) \left(\sqrt{\frac{W(\delta(y/\sigma_X)^2)}{\delta}}\sigma_X\right)'_y \\ &= f_X\left(\sqrt{\frac{W(\delta z^2)}{\delta}}\sigma_X \middle| \boldsymbol{\beta}\right) \sigma_X \frac{1}{\sigma_X} \left(\sqrt{\frac{W(\delta z^2)}{\delta}}\right)'_z \\ &= f_X\left(\sqrt{\frac{W(\delta z^2)}{\delta}}\sigma_X \middle| \boldsymbol{\beta}\right) \frac{1}{2} \left(\frac{W(\delta z^2)}{\delta}\right)^{-\frac{1}{2}} \left(\frac{W(\delta z^2)}{\delta}\right)'_z \\ &= f_X\left(\sqrt{\frac{W(\delta z^2)}{\delta}}\sigma_X \middle| \boldsymbol{\beta}\right) \frac{\sqrt{W(\delta z^2)}}{\sqrt{\delta z(1 + W(\delta z^2))}}, \end{aligned} \quad (8)$$

kus $z = \frac{y}{\sigma_X}$.

Kui me võtame valemities (7) ja (8), et $\sigma_X = 1$, siis saame definitsioonis 4 oleva mittenegatiivselt määratud mittetsentraalse, mitteskaleeritud juhusliku suuruse jaotus- ja tihedusfunktsiooni.

Analoogiliselt saame leida ka asukoha-skaala juhusliku suuruse jaotus- ja tihedusfunktsiooni. Olgu nüüd X asukoha-skaalaperest pärit juhuslik suurus jaotusfunktsiooniga F_X ja tihedusfunktsiooniga f_X ning olgu selle standardhälve σ_X ja keskväärts μ_X . Definitsioonis 6 on defineeritud juhuslik suurus Y raske sabaga Lam-

berti $W \times F_X$ jaotusega ja $U = \frac{X - \mu_X}{\sigma_X}$, nüüd tähistame $z = \frac{y - \mu_X}{\sigma_X}$. Juhusliku suuruse Y jaotusfunktsiooni F_Y saame leida järgnevalt.

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \stackrel{(6)}{=} \mathbb{P}\left(U \exp\left(\frac{\delta}{2}U^2\right) \sigma_X + \mu_X \leq y\right) = \mathbb{P}\left(U \exp\left(\frac{\delta}{2}U^2\right) \leq z\right) \\ &\stackrel{(1)}{=} \mathbb{P}(\delta U^2 \leq W(\delta z^2)) = \mathbb{P}\left(U^2 \leq \frac{W(\delta z^2)}{\delta}\right). \end{aligned}$$

Vaatleme eraldi olukordi $z > 0$ ja $z < 0$. Sarnaselt eelnevale saame $z > 0$ korral, et

$$\begin{aligned} F_Y(y) &= \mathbb{P}\left(U^2 \leq \frac{W(\delta z^2)}{\delta}\right) = \mathbb{P}\left(U \leq \sqrt{\frac{W(\delta z^2)}{\delta}}\right) \\ &= F_X\left(\sqrt{\frac{W(\delta z^2)}{\delta}} \sigma_X + \mu_X\right) \end{aligned}$$

ning $z < 0$ korral saame, et

$$\begin{aligned} F_Y(y) &= \mathbb{P}\left(U^2 \leq \frac{W(\delta z^2)}{\delta}\right) = \mathbb{P}\left(U \leq -\sqrt{\frac{W(\delta z^2)}{\delta}}\right) \\ &= F_X\left(-\sqrt{\frac{W(\delta z^2)}{\delta}} \sigma_X + \mu_X\right). \end{aligned}$$

Seega saime asukoha-skaala raske sabaga Lambert'i $W \times F_X$ juhusliku suuruse jaotusfunktsiooniks

$$F_Y(y|\boldsymbol{\beta}, \delta) = F_X\left(\operatorname{sgn}(z)\sqrt{\frac{W(\delta z^2)}{\delta}} \sigma_X + \mu_X \middle| \boldsymbol{\beta}\right) = F_X(W_\delta(z)\sigma_X + \mu_X|\boldsymbol{\beta}), \quad (9)$$

kus $z = \frac{y - \mu_X}{\sigma_X}$.

Tihedusfunktsiooni kujuks saame valemi (3) põhjal analoogiliselt eelnevale

$$\begin{aligned}
f_Y(y|\boldsymbol{\beta}, \delta) &= f_X(W_\delta(z)\sigma_X + \mu_X|\boldsymbol{\beta}) \sigma_X (W_\delta(z))'_y \\
&= f_X(W_\delta(z)\sigma_X + \mu_X|\boldsymbol{\beta}) (W_\delta(z))'_z \\
&\stackrel{(3)}{=} f_X(W_\delta(z)\sigma_X + \mu_X|\boldsymbol{\beta}) \frac{\text{sgn}(z)\sqrt{W(\delta z^2)}}{\sqrt{\delta}z(1+W(\delta z^2))} \\
&= f_X(W_\delta(z)\sigma_X + \mu_X|\boldsymbol{\beta}) \frac{W_\delta(z)}{z(1+W(\delta z^2))},
\end{aligned} \tag{10}$$

kus $z = \frac{y - \mu_X}{\sigma_X}$.

On selge, kui $\delta = 0$, siis

$$\begin{aligned}
F_Y(y|\boldsymbol{\beta}, \delta) &= \mathbb{P}(Y \leq y|\boldsymbol{\beta}) = \mathbb{P}\left(U \exp\left(\frac{\delta}{2}U^2\right) \sigma_X \leq y \middle| \boldsymbol{\beta}\right) = P(U \leq z|\boldsymbol{\beta}) \\
&= P(X \leq z\sigma_X|\boldsymbol{\beta}) = P(X \leq y|\boldsymbol{\beta}) = F_X(y|\boldsymbol{\beta}),
\end{aligned}$$

kus $z = \frac{y}{\sigma_X}$.

Edasi vaatame lähemalt raske sabaga Lambert'i W juhuslike suuruste kahte olulisemat tõenäosusjaotust.

2.2 Raske sabaga Lambert'i W eksponentjaotus

Olgu X eksponentjaotusega juhuslik suurus parameetriga $\boldsymbol{\beta} = \lambda > 0$ ehk $X \sim \text{Exp}(\lambda)$. Varasemast teame, et juhusliku suuruse X jaotusfunktsioon on

$$F_X(x|\lambda) = 1 - \exp(-\lambda x), \quad \lambda > 0, \quad x \geq 0$$

ja tihedusfunktsioon on

$$f_X(x|\lambda) = \lambda \exp(-\lambda x), \quad \lambda > 0, \quad x \geq 0.$$

Kindlustus- ja finantsvaldkonna rakendusi arvestades pakuvad meile kõige rohkem huvi just positiivselt valitud δ , mille tulemusel muutub eksponentjaotuse parempoolne saba raskemaks. Sel juhul saame paremini kirjeldada kahjude suuruseid. Raske sabaga Lambert'i $W \times \text{Exp}(\lambda)$ jaotuse juhusliku suuruse Y jaotus- ja tihedusfunktsiooni saame leida valemite (7) ja (8) abil. Siin saame kasutada teadmist, et eksponentjaotusega juhusliku suuruse standardhälve on $\sigma_X = \frac{1}{\lambda}$. Meie vaatame ainult juhtu, kui $y \geq 0$. Seega saame juhusliku suuruse Y jaotusfunktsiooniks

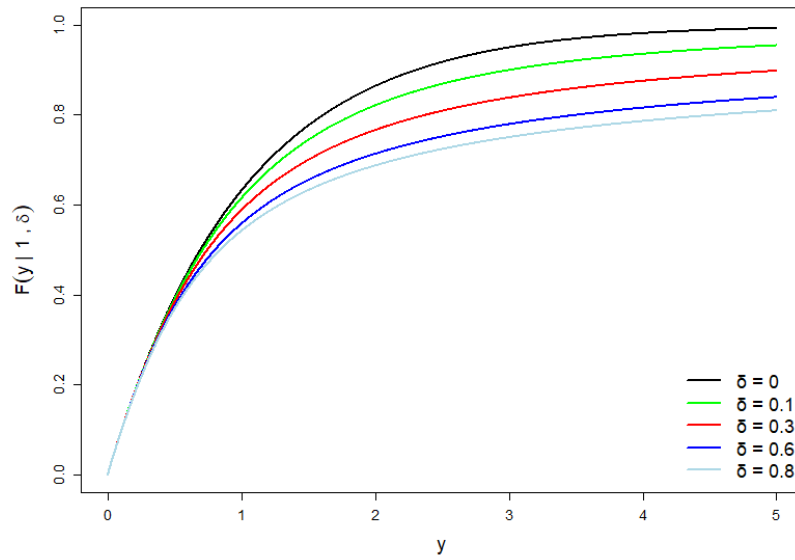
$$\begin{aligned} F_Y(y|\lambda, \delta) &= 1 - \exp\left(-\lambda \sqrt{\frac{W(\delta(y/\sigma_X)^2)}{\delta}} \sigma_X\right) \\ &= 1 - \exp\left(-\sqrt{\frac{W(\delta(y\lambda)^2)}{\delta}}\right), \quad \lambda > 0, \quad y \geq 0 \end{aligned}$$

ja tihedusfunktsiooniks

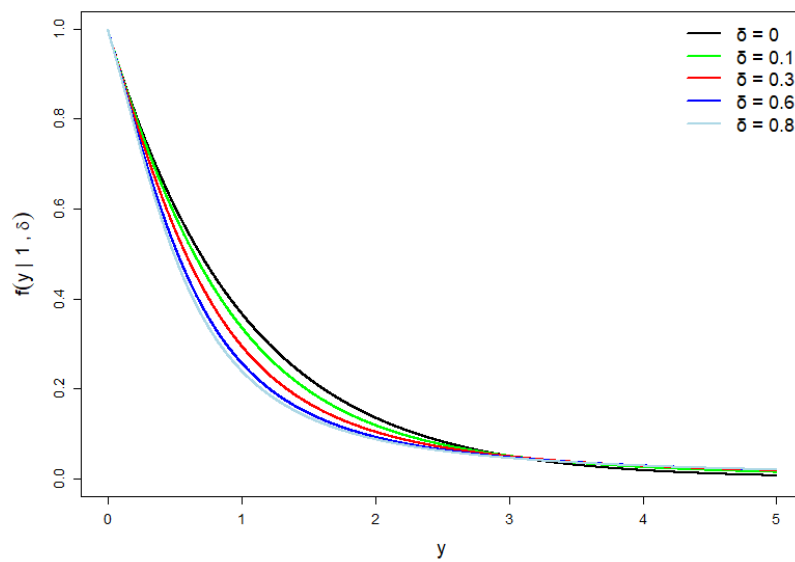
$$\begin{aligned} f_Y(y|\lambda, \delta) &= f_X\left(\sqrt{\frac{W(\delta(y/\sigma_X)^2)}{\delta}} \sigma_X \middle| \beta\right) \frac{\sqrt{W(\delta(y/\sigma_X)^2)}}{\sqrt{\delta} y / \sigma_X (1 + W(\delta(y/\sigma_X)^2))} \\ &= \lambda \exp\left(-\sqrt{\frac{W(\delta(y\lambda)^2)}{\delta}}\right) \frac{\sqrt{W(\delta(y\lambda)^2)/\delta}}{y\lambda(1 + W(\delta(y\lambda)^2))}, \quad \lambda > 0, \quad y \geq 0. \end{aligned}$$

Viimaks paneme need saadud funktsioonid ka joonisele, võtame $\lambda = 1$, siis ka standardhälve $\sigma_X = 1$ ning võtame sabaparametrite väärtusteks $\delta \in \{0; 0,1; 0,3; 0,6; 0,8\}$ (vt joonis 3).

Jooniselt 4 näeme, et eksponentjaotuse tihedusfunktsiooni saba muutub sabaparametri suurenedes raskemaks nii nagu ka Puhkimi magistritöös Lambert'i $W \times \text{Exp}(\lambda)$ jaotuse korral (Puhkim, 2016).



Joonis 3: Raske sabaga Lambert'i $W \times \text{Exp}(1)$ jaotusfunktsioon erinevate sabaparametri väärtuste korral.



Joonis 4: Raske sabaga Lambert'i $W \times \text{Exp}(1)$ tihedusfunktsioon erinevate sabaparametri väärtuste korral.

2.3 Raske sabaga Lambert'i W normaaljaotus

Olgu X normaaljaotusega juhuslik suurus ehk $X \sim N(\mu_X, \sigma_X)$ ning juhusliku suuruse X parameetervektor β on (μ_X, σ_X) .

Sel juhul jaotusfunktsioon F_X avaldub

$$F_X(x|\beta) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right), \quad x \in \mathbb{R},$$

kus Φ on normaaljaotuse jaotusfunktsioon.

Tihedusfunktsioon f_X avaldub kujul

$$f_X(x|\beta) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu_X}{\sigma_X}\right)^2\right), \quad x \in \mathbb{R}.$$

Olgu $\delta > 0$. Selleks, et saaksime leida raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ jaotuse juhusliku suuruse Y jaotus- ja tihedusfunktsiooni, peame kasutama vastavalt valemeid (9) ja (10).

Sel juhul avaldub Y jaotusfunktsioon järgmiselt.

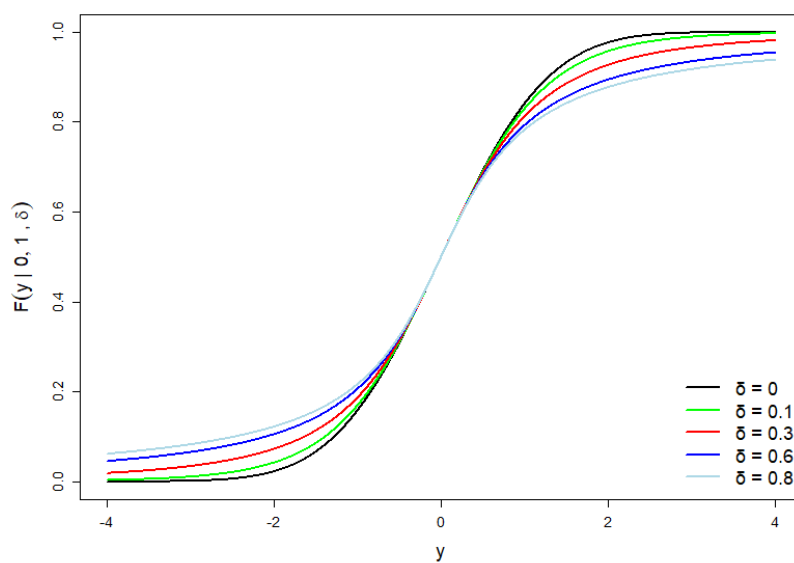
$$F_Y(y|\mu_X, \sigma_X, \delta) = F_X\left(W_\delta\left(\frac{y - \mu_X}{\sigma_X}\right) \sigma_X + \mu_X\right) = \Phi\left(\operatorname{sgn}(z) \sqrt{\frac{W(\delta z^2)}{\delta}}\right), \quad y \in \mathbb{R},$$

kus $z = \frac{y - \mu_X}{\sigma_X}$ ning tihedusfunktsiooni saame

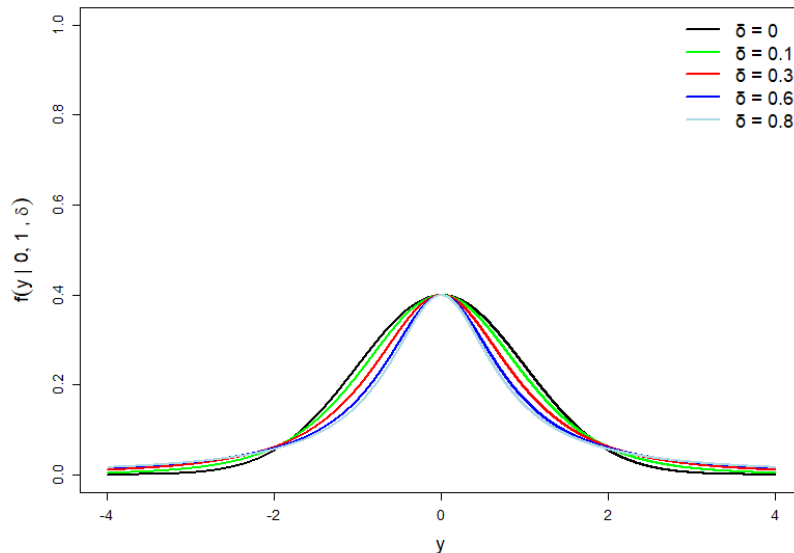
$$\begin{aligned} f_Y(y|\mu_X, \sigma_X, \delta) &= f_X\left(W_\delta\left(\frac{y - \mu_X}{\sigma_X}\right)\right) \cdot \frac{W_\delta((y - \mu_X)/\sigma_X)}{(y - \mu_X)/\sigma_X (1 + W(\delta((y - \mu_X)/\sigma_X)^2))} \\ &= \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left[-\frac{(W_\delta(z))^2}{2}\right] \cdot \frac{W_\delta(z)}{z(1 + W(\delta z^2))}, \\ &\stackrel{(2)}{=} \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left[-\frac{(W_\delta(z))^2}{2}\right] \cdot \exp\left[-\frac{\delta}{2} W_\delta(z)^2\right] \cdot \frac{1}{(1 + W(\delta z^2))} \\ &= \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left[-\frac{1 + \delta}{2} W_\delta(z)^2\right] \cdot \frac{1}{(1 + W(\delta z^2))}, \quad z \in \mathbb{R}, \end{aligned}$$

kus $z = \frac{y - \mu_X}{\sigma_X}$.

Paneme saadud funktsioonid joonisele. Võtame, et $\sigma_X = 1$, $\mu_X = 0$ ja võtame sabaparametrite väärtusteks $\delta = \{0; 0,1; 0,3; 0,6; 0,8\}$. Tihedusfunktsiooni õigsust saame kontrollida sellega, kas tihedusfunktsiooni alune pindala võrdub 1-ga (vt joonis 6).



Joonis 5: Raske sabaga Lambert'i $W \times N(0, 1)$ jaotusfunktsioon erinevate sabaparametrite väärtuste korral.



Joonis 6: Raske sabaga Lambert'i $W \times N(0, 1)$ tihedusfunktsioon erinevate sabaparametrite väärtuste korral.

2.4 Asümmeetriline kahe sabaga juht

Siin alapeatükis kasutame lisaks Goerg'i artiklile (Goerg, 2015b) lisades ära toodud täiendavate materjalide dokumenti (Goerg, 2015a).

Olgu juhuslik suurus Y defineeritud järgmiselt

$$Y = \begin{cases} U \exp\left(\frac{\delta_l}{2} U^2\right), & \text{kui } U \leq 0, \\ U \exp\left(\frac{\delta_r}{2} U^2\right), & \text{kui } U > 0, \end{cases} \quad (11)$$

kus $U \sim N(0, 1)$, $U = \frac{X - \mu_X}{\sigma_X}$ ja $\delta_l, \delta_r \geq 0$. Siin δ_l tähistab juhusliku suuruse Y vasakpoolse saba parameetrit ning δ_r parempoolse saba parameetrit.

Selleks, et leida juhusliku suuruse Y jaotus- ja tihedusfunktsiooni, peame vaatama kahte juhtu, kui $U > 0$ ning $U \leq 0$. See tähendab, et võtame $z = \frac{y - \mu_X}{\sigma_X} > 0$ ja $z = \frac{y - \mu_X}{\sigma_X} \leq 0$, kus μ_X ja σ_X on asukoha-skaala juhusliku suuruse X keskvaartus

ja standardhälve. Seejuures vaatame juhusliku suuruse Y sabaparametrit δ vasakpoole sabaparametri δ_l ja parempoolse sabaparametri δ_r korral eraldi. Seega saame juhusliku suuruse Y jaotusfunktsiooniks

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \stackrel{(11)}{=} \mathbb{P}\left(U \exp\left(\frac{\delta_r}{2}U^2\right) \sigma_X + \mu_X \leq y\right) = \mathbb{P}\left(U \exp\left(\frac{\delta_r}{2}U^2\right) \leq z\right) \\ &\stackrel{(1)}{=} \mathbb{P}(\delta_r U^2 \leq W(\delta_r z^2)) = \mathbb{P}\left(U \leq \sqrt{\frac{W(\delta_r z^2)}{\delta_r}}\right) = F_X\left(\sqrt{\frac{W(\delta_r z^2)}{\delta_r}} \sigma_X + \mu_X\right) \\ &= F_X(W_{\delta_r}(z) \sigma_X + \mu_X), \end{aligned}$$

kui $z > 0$.

Niisiis saime asukoha-skaala raske sabaga Lambert'i $W \times F_X$ juhusliku suuruse parempoolse poole jaotusfunktsiooniks

$$F_Y(y|\boldsymbol{\beta}, \delta_r) = F_X(W_{\delta_r}(z) \sigma_X + \mu_X|\boldsymbol{\beta}), \quad z > 0. \quad (12)$$

Analoogiliselt saame $z \leq 0$ korral juhusliku suuruse Y jaotusfunktsiooni avaldada kujul

$$F_Y(y|\boldsymbol{\beta}, \delta_l) = F_X(W_{\delta_l}(z) \sigma_X + \mu_X|\boldsymbol{\beta}) \quad (13)$$

ja seega oleme saanud tulemuste (12) ja (13) põhjal juhusliku suuruse Y jaotusfunktsiooniks

$$F_Y(y|\boldsymbol{\beta}, \delta_l, \delta_r) = \begin{cases} F_Y(y|\boldsymbol{\beta}, \delta_l), & \text{kui } y \leq 0, \\ F_Y(y|\boldsymbol{\beta}, \delta_r), & \text{kui } y > 0. \end{cases}$$

Juhusliku suuruse Y tihedusfunktsiooniks saame

$$\begin{aligned} f_Y(y|\boldsymbol{\beta}, \delta_r) &= f_X(W_{\delta_r}(z) \sigma_X + \mu_X|\boldsymbol{\beta}) (W_{\delta_r}(z))'_z \\ &\stackrel{(3)}{=} f_X(W_{\delta_r}(z) \sigma_X + \mu_X|\boldsymbol{\beta}) \frac{W_{\delta_r}(z)}{z(1 + W(\delta_r z^2))}, \end{aligned} \quad (14)$$

kus $z > 0$.

Analoogiliselt saame $z \leq 0$ korral juhusliku suuruse tihedusfunktsiooni avaldada kujul

$$f_Y(y|\boldsymbol{\beta}, \delta_l) = f_X(W_{\delta_l}(z)\sigma_X + \mu_X|\boldsymbol{\beta}) \frac{W_{\delta_l}(z)}{z(1 + W(\delta_l z^2))}. \quad (15)$$

Seega oleme saanud tulemuste (14) ja (15) põhjal juhusliku suuruse Y tihedusfunktsiooniks

$$f_Y(y|\boldsymbol{\beta}, \delta_l, \delta_r) = \begin{cases} f_Y(y|\boldsymbol{\beta}, \delta_l), & \text{kui } y \leq 0, \\ f_Y(y|\boldsymbol{\beta}, \delta_r), & \text{kui } y > 0. \end{cases}$$

3 Jaotuste sobitamine andmetele

Käesolevas peatükis keskendume eelneva teooria praktilisele rakendamisele. Analüüsime kahjusummasid, sobitades neile raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ ja raske sabaga Lambert'i $W \times Exp(\lambda)$ jaotust. Töö käigus hindame nende jaotuste sobivust kahjusumma andmetele ning võrdleme tulemusi teiste tuntud jaotustega. Kõigepealt uurime parameetrite hindamist suurima tõepära meetodil, tuginedes Goerg'i artiklile (Goerg, 2015b).

3.1 Parameetrite hindamine

Selleks, et sobitada raske sabaga Lambert'i $W \times F_X$ jaotust andmetele, hindame jaotuse parameetreid suurima tõepära meetodi abil. Vaatame sõltumatute sama jaotusega juhuslike suuruste valimit $(Y_1, \dots, Y_n)$, mille realisatsioon on $(y_1, \dots, y_n)$. Niisiis peame hindama sisendjaotuse parameetrit/parameetreid $\boldsymbol{\beta}$ ja parameetrit δ . Valimi log-tõepärafunktsioon on antud kujul

$$l(\boldsymbol{\beta}, \delta|y) = \sum_{i=1}^N \ln(f_Y(y_i|\boldsymbol{\beta}, \delta)),$$

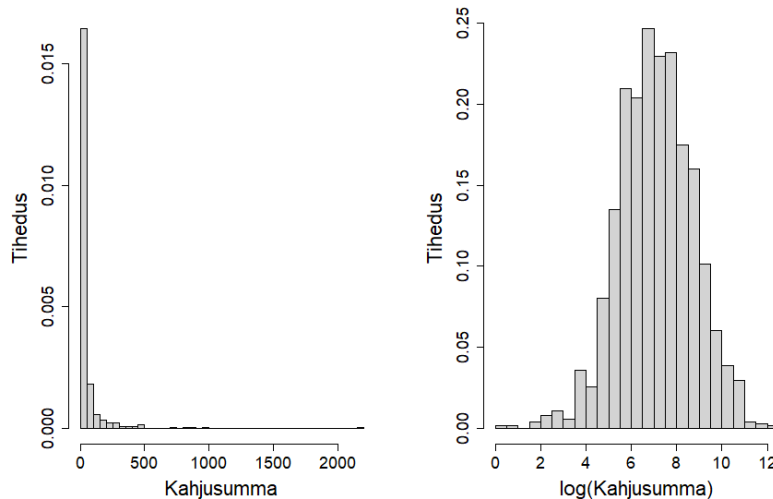
kus raske sabaga Lambert'i $W \times F_X$ jaotuse juhusliku suuruse Y tihedusfunktsioon $f_Y(y|\boldsymbol{\beta}, \delta)$ on antud valemiga (10).

Log-tõepära maksimiseerimisel üle võimalike $(\boldsymbol{\beta}, \delta)$ väärtuste saadakse parameetritele suurima tõepära hinnangud ehk

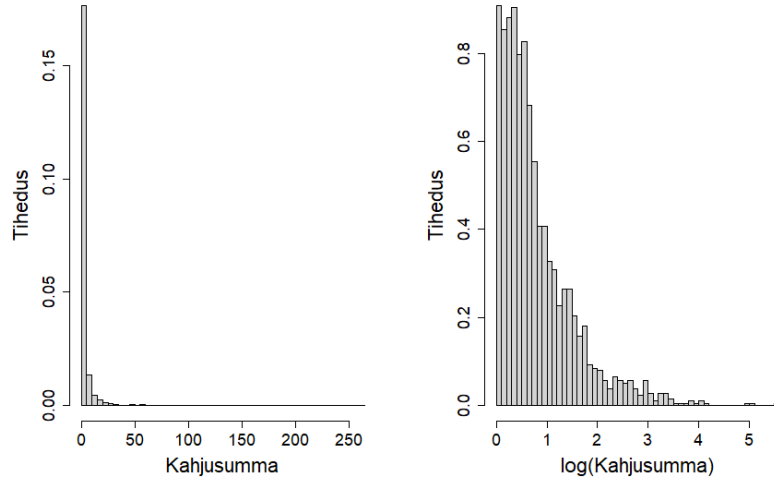
$$(\hat{\boldsymbol{\beta}}, \hat{\delta})_{MLE} = \arg \max_{\boldsymbol{\beta}, \delta} l(\boldsymbol{\beta}, \delta|y).$$

3.2 Jaotuste sobitamine kahjuandmetele

Selles alapeatükis rakendame raske sabaga Lambert'i $W \times F_X$ juhuslike suuruste teooriat kahjuandmetele. Meil on kaks rakendustarkvara R sisseehitatud andmestikku. Esimene andmestik on paketist *copula* nimega *loss* (edaspidi nimega USA andmestik) ning teine andmestik on pärit paketist *fExtremes* nimega *danishClaims* (edaspidi nimega Taani andmestik). Mõlema andmestiku ja paketiga saab lähemalt tutvuda paketimanuaalidest Hofert *et al.*, 2025 ja Wuertz *et al.*, 2023. Esimeses andmestikus on 1 500 kahjusummat. Minimaalne kahju on 10 ja maksimaalne on 2 173 595 USA dollarit. Keskmine kahju suurus on 41 208 ning mediaan on 12 000 USA dollarit. Teises andmestikus on 2 167 kahjusummat, minimaalne on 1,000, maksimaalne 263,250 miljonit Taani krooni. Keskmine kahju on 3,385 ning mediaan on 1,778 miljonit Taani krooni. Selleks, et saaksime tulemusi hiljem võrrelda Kääriku *et al.* saadud Lambert'i $W \times F_X$ jaotuse tulemustega (Käärik *et al.*, 2023) ja Elingi saadud teiste jaotuste tulemustega, teisendame USA andmestiku kujule, kus ühikuks on tuhat USA dollarit (Eling, 2012). Andmestiku logaritmimisel rakendame teisendust $\ln(y) - \min(\ln(y)) + 10^{-10}$ (Käärik *et al.*, 2023).



Joonis 7: USA hüvitiste andmed tuhandetes USA dollarites (vasakul) ning sama andmestik pärast logaritmimist (paremal).



Joonis 8: Taani tulekahjukindlustuse andmed miljonites Taani kroonides (vasakul) ning sama andmestik pärast logaritmimeist (paremal).

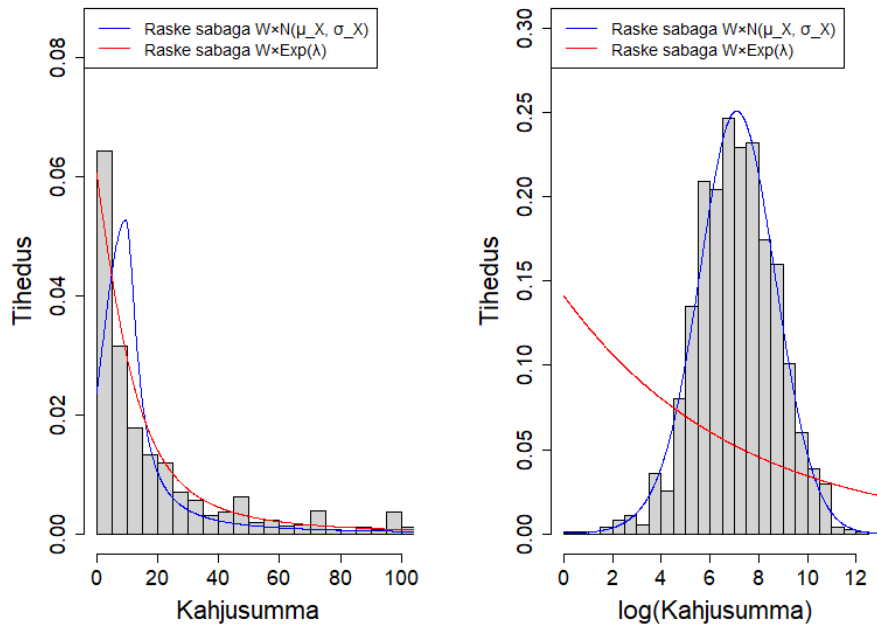
Sobitame nendele andmestikele kahe erineva raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ ja raske sabaga Lambert'i $W \times \text{Exp}(\lambda)$ jaotusi ning võrdleme sobivuse näitajatega Akaike informatsioonikriteeriumi (AIC) põhjal, mis on erinevate jaotuste korral väljatoodud Kääriku *et al.* artiklis (Käärik *et al.*, 2023). Selles artiklis sobitati andmestikele 19 jaotust, millest osutusid AIC näitaja põhjal parimateks USA andmestikul logaritmiline normaaljaotus ja logaritmitud USA andmestikul asümmeetriline normaaljaotus ning Lambert'i W normaaljaotus. Taani andmestikul osutus parimaks asümmeetriline t-jaotus ning logaritmitud Taani andmestikul Lambert'i W normaaljaotus (Käärik *et al.*, 2023).

Kõigepealt toome välja hinnatud parameetrid raske sabaga Lambert'i W jaotuste puhul.

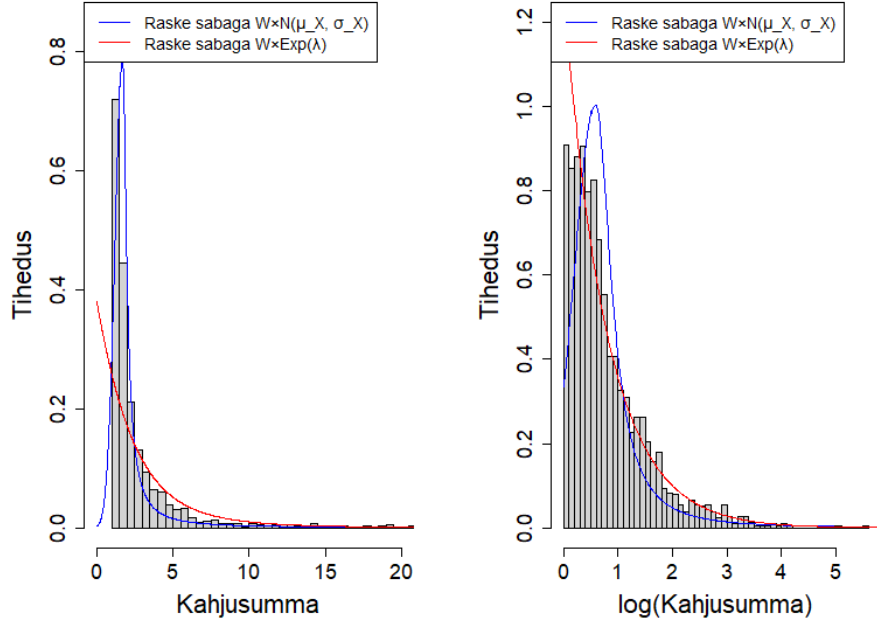
Tabel 1: Hinnatud parameetrid raske sabaga Lambert'i W jaotuste jaoks.

	$W \times \text{Exp}(\lambda)$		$W \times N(\mu_X, \sigma_X)$			
	λ	δ	μ_X	σ_X	δ_l	δ_r
USA andmestik	0,060	0,286	9,575	7,562	$1,511e^{-8}$	1,857
USA log andmestik	1,414	$9,822e^{-10}$	7,092	1,591	0,036	$1,210e^{-8}$
Taani andmestik	0,381	0,058	1,705	0,507	$7,863e^{-16}$	1,585
Taani log andmestik	1,270	$1,365e^{-15}$	0,596	0,402	$6,814e^{-9}$	0,567

Visualiseerime saadud jaotusi ka joonistel koos mainitud andmete histogrammidega (vt jooniseid 9 ja 10). Sarnaselt Kääriku *et al.* artiklis olevatele joonistele näitame histogrammil USA andmestiku puhul kahjuandmeid, mis on alla 100 tuhande USA dollari ja Taani andmestiku puhul kuvame kahjuandmeid, mis on alla 20 miljoni Taani krooni (Käärik *et al.*, 2023).



Joonis 9: USA andmete sobitatud jaotused (vasakul) ning logaritmitud USA andmestikule sobitatud jaotused (paremal).



Joonis 10: Taani andmete sobitatud jaotused (vasakul) ning logaritmitud Taani andmestikule sobitatud jaotused (paremal).

Jooniselt 9 selgub, et raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ jaotus võiks kirjeldada logaritmitud USA andmeid hästi. USA andmestiku puhul on näha, et raske sabaga Lambert'i $W \times \text{Exp}(\lambda)$ võiks sobida andmetele tunduvalt paremini kui raske sabaga Lambert'i W normaaljaotus. Samuti joonise 10 põhjal võiks arvata, et logaritmitud Taani andmetele sobib raske sabaga Lambert'i W eksponentjaotus hästi.

Meie eesmärk on võrrelda, kas raske sabaga Lambert'i W jaotused sobivad andmetele paremini kui Lambert'i W jaotused. Võrdleme tulemusi Kääriku *et al.* artiklis saadud AIC tulemustega (Käärik *et al.*, 2023). Tabelist 2 näeme, et ainult logaritmitud andmestikul osutus raske sabaga Lambert'i W normaaljaotus paremaks ja algandmestikul oli raske sabaga Lambert'i W eksponentjaotus lähedane Lambert'i W eksponentjaotuse omaga, seda kinnitavad saadud AIC väärtused. Andmete iseloomu põhjal oli oodata, et raske sabaga Lambert'i $W \times \text{Exp}(\lambda)$ sobib USA logaritmitud andmestikule halvemini, sest see on sarnane normaaljaotuse ku-

juga ning AIC kriteerium kinnitab seda. Taani andmestiku puhul osutus paremaks algandmestikul raske sabaga Lambert'i W eksponentjaotus ja logaritmitud andmestikul raske sabaga Lambert'i W normaaljaotus. Üsna ligilähedane tulemus tuli ka raske sabaga Lambert'i W eksponentjaotuse AIC sobivusnäitaja logaritmitud andmestikul (vt Tabel 3).

Tabel 2: USA andmestikele sobitatud jaotuste AIC sobivusnäitajad.

AIC		
Jaotus	Algandmed	Logaritmitud
Lambert'i $W \times Exp(\lambda)$	13 141,92	7 845,81
Lambert'i $W \times N(\mu_X, \sigma_X)$	13 397,48	5 737,79
Raske sabaga Lambert'i $W \times Exp(\lambda)$	13 209,29	8 871,95
Raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$	14 002,10	5 734,95

Tabel 3: Taani andmestikele sobitatud jaotuste AIC sobivusnäitajad.

AIC		
Jaotus	Algandmed	Logaritmitud
Lambert'i $W \times Exp(\lambda)$	9 264,10	3 282,22
Lambert'i $W \times N(\mu_X, \sigma_X)$	6 699,82	2 978,46
Raske sabaga Lambert'i $W \times Exp(\lambda)$	9 111,25	3 295,77
Raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$	7 546,88	3 944,75

Lisades A.1 ja A.2 on artiklitest Käärrik *et al.*, 2023 ja Eling (2012) saadud AIC väärtused koos raske sabaga Lambert'i W jaotuste AIC kriteeriumitega. Näeme, et USA algandmestiku puhul esikolmik ei muutunud ning raske sabaga Lambert'i $W \times Exp(\lambda)$ jaotus on paremuselt seitsmendal kohal. AIC kriteeriumi põhjal sobis andmetega kõige paremini raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ jaotus. Taani algandmestiku puhul samuti esikolmik ei muutunud ja raske sabaga Lambert'i W jaotuste sobivusnäitajate tulemused jäävad pigem tahapoole. Taani logaritmitud andmete korral esikolmik ei muutunud, kuid tulemuse poolest ei jää raske sabaga Lambert'i $W \times Exp(\lambda)$ jaotus (AIC 3 295,77) teistele väga palju alla.

Kokkuvõte

Antud bakalaureusetöö on jätk Puhkimi magistritööle (Puhkim, 2016). Töös anti põgus ülevaade Lambert'i W funktsioonist ning abitulemustest, mida hilisemas peatükis kasutati. Lisaks toodi välja skaalapere ja asukoha-skaalapere definitsioonid koos lihtsate näidetega ning tutvustati raske sabaga Lambert'i W juhuslikke suursusi. Seejuures tuletati nende jaotus- ja tihedusfunktsiooni üldkuju.

Lühidalt analüüsiti olukorda, kui sisendiks oli eksponentjaotus ja normaaljaotus ning anti ülevaade asümmeetrilisest kahe sabaga juhust ehk kui juhuslik suurus on defineeritud vasak- ja parempoolse saba kaudu. Töö praktilises osas rakendati raske sabaga Lambert'i $W \times \text{Exp}(\lambda)$ ja kahe erineva raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ juhtu R sisseehitatud kahe algandmestiku peal paketest *fExtremes* nimega *danishClaims* (Wuertz *et al.*, 2023) ja paketest *copula* nimega *loss* (Hofert *et al.*, 2025) ning nende logaritmitud andmete peal. Suurima tõepära meetodiga leiti parameetritele hinnangud, seejärel analüüsiti nende jaotuste sobivust mainitud neljal andmestikul. AIC sobivusnäitajaid võrreldi varasemalt saadud Kääriku *et al.* (Käärik *et al.*, 2023) ja Elingi artiklites olevate tulemustega (Eling, 2012). Lisaks Lambert'i W jaotustele võrreldi tulemusi artiklis väljatoodud 19 kahjujaotusega. Osutus, et USA logaritmitud andmestikule sobis kõigist jaotustest kõige paremini raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ jaotus, kuid ülejäänud andmestike puhul jäid raske sabaga Lambert'i W jaotused Lambert'i W jaotustele ja teistele tuntud kahjujaotustele alla. Taani andmestiku puhul oli raske sabaga Lambert'i $W \times \text{Exp}(\lambda)$ jaotus Lambert'i $W \times \text{Exp}(\lambda)$ jaotusest parem, kuid siiski jäi teistele kahjujaotustele alla.

Raske sabaga Lambert'i W juhuslikud suurused on vähe uuritud ja paindlik võimalus kahjuandmete kirjeldamiseks. Saadud tulemuste põhjal raske sabaga Lambert'i $W \times \text{Exp}(\lambda)$ jaotus sobis enamjaolt hästi. Raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$ jaotuse jaoks ei olnud andmed kõige paremad, kuna positiivse δ -ga ei saa teha vasakut saba kergemaks kui normaaljaotusel. Edasiseks arenduseks, kui tarkvaraliselt

oleks võimalus sobitada negatiivset vasakpoolse saba parameetrit, siis oleks ilmselt ka praeguste andmetega parema sobivuse leidnud.

Kasutatud allikad

- Casella, G. ja R. L. Berger (2001). *Statistical Inference. Location and Scale Families*. 2. väljaanne. Belmont: Carolyn Crockett, lk. 116–121. URL: https://pages.stat.wisc.edu/~shao/stat610/Casella_Berger_Statistical_Inference.pdf.
- Delignette-Muller, M.-L., C. Dutang, R. Pouillot, J.-B. Denis ja A. Siberehicot (7. jaanuar 2025). *Help to Fit of a Parametric Distribution to Non-Censored or Censored Data*. Versioon 1.2-2. URL: <https://cran.r-project.org/web/packages/fitdistrplus/fitdistrplus.pdf> (vaadatud 30.04.2025).
- Eling, M. (2012). „Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?“ *Science Direct*. URL: https://www.sciencedirect.com/science/article/pii/S0167668712000492?casa_token=eLSnE-OymrcAAAAA:xJRuolob7nRTt0IIeRyWUk4-0tMhYHH5viZA69wrpONZKP4g7u_1wyQHVPDCJa1dRDrxuG-pz8Q (vaadatud 05.05.2025).
- Goerg, G. M. (2015a). „Supplementary Material of “The Lambert Way to Gaussianize heavy tailed data with the inverse of Tukey’s h transformation as a special case”“. *The Scientific World Journal*. URL: <https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1155%2F2015%2F909231&file=tswj909231-sup-0001-f1.pdf> (vaadatud 23.04.2025).
- Goerg, G. M. (2015b). „The Lambert Way to Gaussianize Heavy-Tailed Data with the Inverse of Tukey’s h Transformation as a Special Case“. *The Scientific World Journal*. URL: <https://onlinelibrary.wiley.com/doi/10.1155/2015/909231> (vaadatud 23.04.2025).

- Goerg, G. M. (2025). *Probabilistic Models to Analyze and Gaussianize Heavy-Tailed, Skewed Data*. Versioon 0.6.9-1. URL: <https://cran.r-project.org/web/packages/LambertW/LambertW.pdf> (vaadatud 27.04.2025).
- Hofert, M., I. Kojadinovic, M. Maechler, J. Yan, J. G. Nešlehova ja R. Morger (18. märts 2025). *Multivariate Dependence with Copulas*. Versioon 1.1-6. URL: <https://cran.r-project.org/web/packages/copula/copula.pdf> (vaadatud 27.04.2025).
- Käärrik, M., A. Selart, T. Puhkim ja L. Tee (2023). „Lambert W Random Variables and Their Applications in Loss Modelling“. *MDPI*. URL: <https://www.mdpi.com/2073-8994/15/10/1877> (vaadatud 23.04.2025).
- Puhkim, T. (2016). „Lambert'i W juhuslikud suurused ja nende rakendamine kahjukindlustuses“. Magistritöö. Tartu: Tartu Ülikool. URL: <https://dspace.ut.ee/server/api/core/bitstreams/41fec10c-ff8e-4eff-8f04-aaa1fdd03173/content>.
- Wuertz, Diethelm *et al.* (21. detsember 2023). *Rmetrics - Modelling Extreme Events in Finance*. Versioon 4032.84. URL: <https://cran.r-project.org/web/packages/fExtremes/fExtremes.pdf> (vaadatud 27.04.2025).

Lisa A Sobitatud jaotuste AIC sobivusnäitajad

A.1 USA andmestikele sobitatud 21 jaotuse AIC sobivusnäitajad

Tabel 4: USA andmestikele sobitatud 21 jaotuse AIC sobivusnäitajad.

AIC		
Jaotus	Algandmed	Logaritmitud
Lambert'i $W \times Exp(\lambda)$	13 141,92	7 845,81
Lambert'i $W \times N(\mu_X, \sigma_X)$	13 397,48	5 737,79
Raske sabaga Lambert'i $W \times Exp(\lambda)$	13 209,29	8 871,95
Raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$	14 002,10	5 734,95
EkspONENT-	14 157,93	8 869,95
GAMMA-	13 537,17	6 442,22
Log-normaalne	13 137,53	8 895,12
Logistiline	16 544,91	5 753,92
Normaal-	18 156,65	5 740,44
Weibull	13 321,70	5 923,95
Cauchy	14 518,07	6 264,44
Pareto	13 148,51	8 871,95
Sümmeetriline hüperboolne	15 884,38	5 738,41
Sümmeetriline normaalne Gaussi pöörd-	14 515,76	5 738,38
Sümmeetriline dispersiooni-gamma- ¹	14 261,53	5 738,65
Studenti t-	14 492,64	5 738,12
Asümmeetriline normaal.	16 315,13	5 737,79
Asümmeetriline hüperboolne	14 163,24	5 738,16
Asümmeetriline normaalne Gaussi pöörd-	13 148,66	5 738,12
Asümmeetriline dispersiooni-gamma-	14 177,46	5 738,61
Sümmeetriline üldistatud hüperboolne	14 494,64	5 740,43
Asümmeetriline t-	13 197,79	5 738,06
Asümmeetriline üldistatud hüperboolne	<u>13 145,91</u>	5 740,61

¹ ingl *variance gamma*

A.2 Taani andmestikele sobitatud 21 jaotuse AIC sobivusnäitajad

Tabel 5: Taani andmestikele sobitatud 21 jaotuse AIC sobivusnäitajad.

AIC		
Jaotus	Algandmed	Logaritmitud
Lambert'i $W \times Exp(\lambda)$	9 264,10	3 282,22
Lambert'i $W \times N(\mu_X, \sigma_X)$	6 699,82	2 978,46
Raske sabaga Lambert'i $W \times Exp(\lambda)$	9 111,25	3 295,77
Raske sabaga Lambert'i $W \times N(\mu_X, \sigma_X)$	7 546,88	3 944,75
EkspONENT-	9 620,79	3 297,61
GAMMA-	9 538,19	3 299,61
Log-normaalne	8 119,79	5 504,62
Logistiline	11 479,71	4 421,17
Normaal-	15 431,52	4 709,15
Weibull	9 611,24	3 294,27
Cauchy	8 240,17	4 589,38
Pareto	9 249,67	3 318,07
Sümmeetriline hüperboolne	10 433,17	4 363,90
Sümmeetriline normaalne Gaussi pöörd-	8 237,61	4 303,93
Sümmeetriline dispersiooni-gamma- ¹	9 089,69	4 375,17
Studenti t-	8 237,85	4 299,90
Asümmeetriline normaal.	12 608,36	3 441,49
Asümmeetriline hüperboolne	8 109,27	3 307,83
Asümmeetriline normaalne Gaussi pöörd-	6 806,79	3 378,14
Asümmeetriline dispersiooni-gamma-	7 404,07	3 281,06
Sümmeetriline üldistatud hüperboolne	8 224,65	4 298,21
Asümmeetriline t-	6 683,02	3 274,24
Asümmeetriline üldistatud hüperboolne	6 775,85	3 283,06

¹ ingl *variance gamma*

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Anet Ilustrumm,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Raske sabaga Lambert’i W teisendus ja selle kasutamine kahjujaotuste hindamisel“, mille juhendajad on Meelis Käärik ja Anne Selart, 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 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.

Anet Ilustrumm

14.05.2025