

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Dora Stern
**Kaheosalise normaaljaotuse konstrueerimine ja
selle sobitamine andmetele**

Matemaatilise statistika eriala
Bakalaureusetöö (9 EAP)

Juhendaja: Anne Selart, MSc

TARTU 2026

KAHEOSALISE NORMAALJAOTUSE KONSTRUEERIMINE JA SELLE SOBITAMINE ANDMETELE

Bakalaureusetöö

Dora Stern

Lühikokkuvõte

Antud bakalaureusetöö eesmärk on tutvustada kaheosaliste jaotuste meetodit, keskendudes kaheosalisele normaaljaotusele. Teoreetilises osas tuletakse kaheosalise normaaljaotuse tihedus-, jaotus- ja kvantiilfunktsioon ning leitakse momendid, asümmeetria- ja järsakuskordaja. Praktilises osas hinnatakse kaheosalise normaaljaotuse parameetrid genereeritud ja reaalsele andmetele. Mõlemate juhul vaadeldakse jaotuse sobivust andmetele võrreldes teiste jaotuste AIC näitajaga. Töö teoreetilise osa aluseks on artikkel “A Simple Skewed Distribution with Asset Pricing Applications” (Roon ja Karehnke, [2017](#)).

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

Märksõnad: *kaheosaline jaotus, normaaljaotus, asümmeetrilised andmed*

CONSTRUCTION AND FITTING OF THE TWO-PIECE NORMAL DISTRIBUTION

Bachelor thesis

Dora Stern

Abstract

The aim of this bachelor's thesis is to introduce the method of two-piece distributions, focusing specifically on the two-piece normal distribution. In the theoretical part, the most important functions of the two-piece normal distribution are constructed, and its moments, skewness, and kurtosis are derived. In the practical part, the parameters of the two-piece normal distribution are estimated and fitted to both simulated and real data. In both cases, the goodness-of-fit is compared to other distributions using AIC values. The theoretical part of the thesis is based on the article "A Simple Skewed Distribution with Asset Pricing Applications" (Roon ja Karehnke, [2017](#)).

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

Key Words: *two-piece distribution, normal distribution, split-normal, skewed data*

Sisukord

Sissejuhatus	5
1 Kaheosaliste jaotuste meetod	7
1.1 Tutvustus	7
1.2 Metoodika	8
2 Kaheosaline normaaljaotus	11
2.1 Eelteadmised	11
2.2 Tihedusfunktsioon, jaotusfunktsioon ja kvantiilid	12
2.3 Momentide leidmine	16
2.4 Asümmeetria- ning järsakuskordaja	22
2.5 Numbriline näide	24
3 Jaotuse sobitamine andmetele	27
3.1 Parameetrite hindamine	27
3.2 Jaotuse sobitamine genereeritud andmetele	28
3.3 Jaotuse sobitamine reaalsele kahjuandmetele	30
Kokkuvõte	33
Kasutatud allikad	35
Lisa A: Kaheosalise normaaljaotusega andmete genereerimine ja parameetrite hindamine	37

**Lisa B: Logaritmitud kahjuandmete sobitatud 20 jaotuse AIC
sobivusnäitajad**

39

Sissejuhatus

Andmeanalüüsis on oluliseks märksõnaks jaotus, mis kirjeldab andmete paiknevust ja hajuvust mõne arvkarakteristiku suhtes. Normaaljaotus on laialdaselt kasutuses ning tuntud oma kellukese kuju ja heade teoreetiliste omaduste poolest. Praktikas võib see aga anda eksitavaid tulemusi, kui päris andmed ei vasta normaaljaotuse poolt eeldatud sümmeetria- ja kergetele sabadele. Andmete asümmeetriaga arvestamiseks on koostatud mitmeid meetodeid ja üks neist on kaheosalise normaaljaotuse (ingl *two-piece normal distribution*/*split-normal distribution*) meetod. Selle tööpõhimõtteks on võtta kahest erineva hajuvuse, kuid sama keskväärtusega normaaljaotusest pool ning skaaleerida need pooled pidevaks jaotuseks kokku.

Bakalaureusetöö eesmärgiks on tutvustada kaheosaliste jaotuste (ingl *two-piece distribution*) meetodit, keskendudes kaheosalisele normaaljaotusele. Anda ülevaade jaotuse toimimisest, matemaatilisest taustast ja sobitada seda ka andmetele.

Käesolev töö koosneb kolmest suuremast peatükist. Esimeses peatükis antakse ülevaade kaheosaliste jaotuste meetodi ideest ning ajaloost. Lisaks tutvustatakse kaheosalise tihedusfunktsiooni üldkuju ning defineeritakse jaotuse kirjeldamiseks vajalikud momendid, asümmeetriakordaja (ingl *skewness*) ja järsakuskordaja (ingl *kurtosis*).

Teises peatükis keskendutakse kaheosalise normaaljaotuse teoreetilisele taustale, tuletatakse selle tihedus-, jaotus- ja kvantiilfunktsiooni. Samuti leitakse kaheosalise normaaljaotuse teoreetilised momendid, asümmeetria- ja järsakuskordaja ning nende piirid. Tulemused põhinevad suuresti artiklil “A Simple Skewed Distribution with Asset Pricing Applications” (Roon ja Kahrenke, 2017). Tehakse ka numbriline näide, et demonstreerida parameetrite,

momentide ja kordajate tõlgendamist.

Kolmandas peatükis tutvustatakse kaheosalise normaaljaotuse parameetrite hindamist ning rakendatakse seda genereeritud andmetel. Seejärel sobitatakse kaheosalist normaaljaotust ka reaalsele kahjuandmetele. Jaotus sobivust andmetega võrreldakse Akaike informatsioonikriteeriumiga (AIC).

Jooniste koostamiseks, arvutuste tegemiseks ning praktilise osa läbiviimiseks on kasutatud statistikatarkvara R (R Core Team, [2025](#)). Allikate ja andmestike otsimiseks ning R-i koodi kirjutamiseks on kasutatud tehisintellekti abi (Microsoft, [2026](#)). Töö on kirjutatud ja vormistatud Overleaf keskkonnas (Overleaf, [2026](#)).

1 Kaheosaliste jaotuste meetod

Selles peatükis anname ülevaate kaheosaliste jaotuste meetodi toimimisest, taustast ning meetodi kasutamiseks vajalikest eelteadmistest.

1.1 Tutvustus

Jaotus näitab seda, kuidas andmed mõne oma arvkarakteristiku ümber paiknevad. Näiteks normaaljaotuse puhul on tegemist sümmeetrilise kellukesekujulise jaotusega. See tähendab, et enim väärtuseid asub jaotuse keskpunktis ning mõlemale poole seda punkti jääb võrdse hajuvusega järjest vähem väärtusi (sabad) ehk äärmuslikke tulemusi esineb kõige harvem. Normaaljaotus töötab teoreetiliselt, kuid reaalse andmete puhul ei esine alati sümmeetrisust ning jaotuse sabad ei pruugi olla võrdse raskusega. Sellises olukorras võib toimida kaheosaliste jaotuste meetod. (Rubio ja Steel, [2020](#))

Kaheosalised jaotused said laiemalt tuntuks umbes 30 aastat tagasi, kuid tegelikult pärineb see meetod 19. sajandi lõpust. Esimest korda mainiti kaheosalist normaaljaotust G. T. Fechneri 1897. aasta teoses “Kollektivmasslehre”, kus Fechner esitas “Zweiseitige Gauss’sche Gesetz” ehk kahepoolse Gaussi seaduse, mis lubab panna jaotuse vasakule ja paremale poolele erineva dispersiooni. Fechneri autorlust on hiljem kahtluse alla seadnud K. Pearson, kes väitis, et sarnaseid ideid ilmus teiste autorite poolt juba varemgi. K. F. Wallis aga tõestas, et Pearsoni spekulatsioonid ei olnud õiglased ning Fechneri panust meetodi loomisel alahinnati. Pärast Fechnerit on kaheosaliste jaotuste ideed edasi arendanud näiteks F. Y. Edgeworth, hiljem on meetodit uurinud veel mitmed statistikud ja ka füüsikud. (Wallis, [2014](#))

1.2 Metoodika

Kaheosalise jaotuse loomise eelduseks on sümmeetriline tihedusfunktsioon f , millel on asukohaparameter μ , skaalaparameter σ ja kujuparameter δ . Baasjaotuse õige valik on väga oluline, sest see mõjutab sabade kuju ning sobivust andmetega. Murdes baasjaotuse keskelt pooleks ning skaleerides pooled erinevalt, saamegi kaheosalise jaotuse. Pärast ühendatakse erinevate hajuvustega pooled taas kokku üheks tihedusfunktsiooniks h , mille võib üldkujul kirja panna järgmiselt:

$$h(x; \mu, \sigma_1, \sigma_2, \delta) = \begin{cases} \frac{2}{\sigma_1 + \sigma_2} f\left(\frac{x - \mu}{\sigma_1}; \delta\right), & \text{kui } x \leq \mu, \\ \frac{2}{\sigma_1 + \sigma_2} f\left(\frac{x - \mu}{\sigma_2}; \delta\right), & \text{kui } x > \mu. \end{cases}$$

Siin σ_1 ja σ_2 on skaalaparameetrid vastavalt vasaku ja parema poole jaoks ning δ kirjeldab sabade kuju. Skaalaparameetrite vahelist suhet saab kasutada sabade kaalumiseks ehk skaleerimiseks, et tihedus oleks õigesti normeeritud. Selline meetod lisab modelleerimisele palju paindlikkust, sest võimaldab jaotust andmetega paremini sobitada. (Rubio ja Steel, 2020)

Kaheosalise jaotuse koostamisel on eriti oluline kahe kombineeritava jaotuse erinev hajuvus, mis põhjustab asümmeetria. Saadud asümmeetrilise jaotuse kuju kirjeldamiseks kasutatakse näiteks asümmeetriakordajat, mis näitab jaotuse kallutatust ja järsakuskordajat, mis näitab sabade raskust. (Rubio, 2013)

Selleks, et eelnimetatud kordajad leida, kasutame momente.

Definitsioon 1. Juhusliku suuruse X k -ndat järku momendiks nimetame arve $m_k = E(X)^k$, kus k on positiivne täisarv.

Seejuures märgime, et esimest järku moment on võrdne keskvärtusega EX

ning teist järku momendi kaudu saab leida dispersiooni $DX = m_2 - m_1^2$. Pideva juhusliku suuruse X k -ndat järku momendi leidmiseks kasutame järgmist seost:

$$m_k = E(X)^k = \int_{-\infty}^{\infty} x^k \cdot f(x) dx \quad (1)$$

Edaspidi rakendame ka tsentraalmomente, mis mõõdavad hajuvust keskmise ümber.

Definitsioon 2. Juhusliku suuruse X k -ndat järku tsentraalmomendiks nimetame arve $E[(X - EX)^k]$, kus k on positiivne täisarv.

Asümmeetriakordaja leiame kolmandat järku tsentraalmomendi abil:

$$\begin{aligned} E[(X - EX)^3] &= E[X^3 - 3X^2EX + 3X(EX)^2 - (EX)^3] \\ &= E(X)^3 - (EX)^3 - 3EX[E(X)^2 - (EX)^2] \\ &= E(X)^3 - (EX)^3 - 3EXDX. \end{aligned} \quad (2)$$

Standardiseerimata kordaja puhul võib andmete ühikute erinevus arvutusi segada, seega jagame tulemuse standardhälbega, et kordaja standardiseerida, ning edaspidi kasutame alljärgnevat definitsiooni.

Definitsioon 3. Juhusliku suuruse X jaotuse asümmeetriakordajaks nimetame suurust

$$\gamma_1 = E\left[\left(\frac{X - EX}{\sqrt{DX}}\right)^3\right] = \frac{E[(X - EX)^3]}{(\sqrt{DX})^3}. \quad (3)$$

Järsakuskordaja leiame analoogselt standardiseeritud neljandat järku tsentraalmomendist:

$$\begin{aligned}
E[(X - EX)^4] &= E[X^4 - 4X^3EX + 6X^2(EX)^2 - 4X(EX)^3 + (EX)^4] \\
&= E(X)^4 - 4EXE(X)^3 + 6(EX)^2E(X)^2 - 4(EX)^3EX + (EX)^4 \\
&= E(X)^4 - 4EXE(X)^3 + 6(EX)^2E(X)^2 - 3(EX)^4. \tag{4}
\end{aligned}$$

Definitsioon 4. Juhusliku suruse X jaotuse järsakuskordajaks nimetame suurust

$$\gamma_2 = E \left[\left(\frac{X - EX}{\sqrt{DX}} \right)^4 \right] = \frac{E[(X - EX)^4]}{(\sqrt{DX})^4}. \tag{5}$$

2 Kaheosaline normaaljaotus

Järgnevas peatükis tutvustame kaheosaliste jaotuste meetodit teoreetiliselt, normaaljaotuse põhjal. Konstrueerime kaheosalise normaaljaotuse tihedusfunktsiooni, jaotusfunktsiooni ja kvantiilfunktsiooni ning leiame momendid, tuginedes Rooni ja Karehnke'i artiklis "A Simple Skewed Distribution with Asset Pricing Applications" esitatud tulemusele (Roon ja Karehnke, 2017).

2.1 Eelteadmised

Tuletame meelde normaaljaotuse tihedusfunktsiooni. Kui juhuslik suurus X on normaaljaotusega $X \sim N(\mu, \sigma^2)$, kus μ on asukohaparameter ehk antud juhul keskvärtus ning σ on skaalaparameter ehk standardhälve, siis kirjutame tema tihedusfunktsiooni järgnevalt:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (6)$$

Märgime ka ära, et normaaljaotuse jaotusfunktsioon pannakse tavaliselt kirja standardse normaaljaotuse $Z \sim N(0, 1)$ jaotusfunktsiooni $\Phi(z)$ kaudu järgnevalt:

$$F(x; \mu, \sigma) = \int_{-\infty}^x f(x; \mu, \sigma) dx = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad (7)$$

Hiljem kasutame järgmist teadmist normaaljaotuse absoluutse tsentraalmomendi kohta:

$$E[|X - \mu|^k] = \sigma^k (k-1)!! \times \begin{cases} \sqrt{2/\pi}, & \text{kui } k \text{ on paaritu,} \\ 1, & \text{kui } k \text{ on paaris.} \end{cases} \quad (8)$$

Siin k näitab mitmendat järku momendiga on tegu ning $(k-1)!!$ tähistab

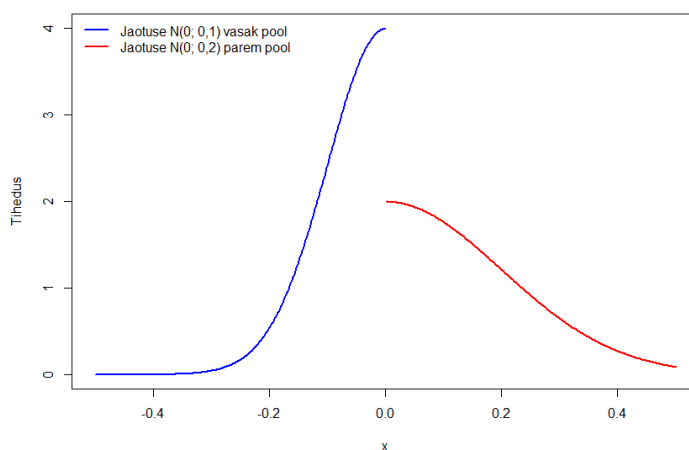
sama paarsusega arvude korrutist kuni arvuni $k - 1$. (Winkelbauer, 2014)

2.2 Tihedusfunktsioon, jaotusfunktsioon ja kvantiilid

Kirjeldame kaheosalise normaaljaotusega juhusliku suuruse X tihedusfunktsiooni esialgu järgnevalt

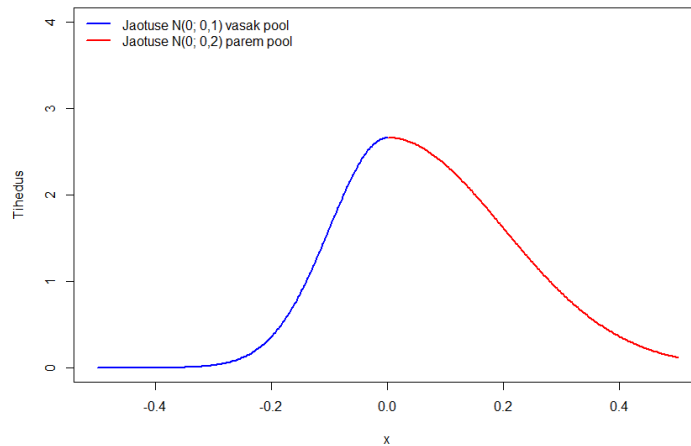
$$g_*(x; \mu, \sigma_1, \sigma_2) = \begin{cases} f(x; \mu, \sigma_1), & \text{kui } x \leq \mu, \\ f(x; \mu, \sigma_2), & \text{kui } x > \mu. \end{cases}$$

Siin f tähistab normaaljaotuste $N(\mu, \sigma^2)$ tihedusfunktsiooni, μ keskväärtust ning σ_1 ja σ_2 on vastavalt esimese ja teise normaaljaotuse standardhälbed. Olgugi, et g_* koosneb kahest sümmeetrilisest jaotusest, siis kui nende standardhälbed omavahel ei võrdu, on tegemist asümmeetrilise jaotusega.



Joonis 1: Kaks poolikut normaaljaotuse tihedusfunktsiooni

Sellisel kujul määratud tihedusfunktsioon g_* ei ole pidev punktis $x = \mu$, sest kahe poole tihedused ei ühti (Joonis 1). Seetõttu skaleerime kummagi poole vastavalt teguritega λ_1 ja λ_2 (Joonis 2).



Joonis 2: Kaks normaaljaotust skaleeritud pidevaks tihedusfunktsiooniks

Pidevuse jaoks nõuame, et keskväärtuse juures kahe normaaljaotuse tihedusfunktsioonid ühtiksid

$$\lambda_1 f(\mu; \mu, \sigma_1) = \lambda_2 f(\mu; \mu, \sigma_2).$$

Nõuame, et kohal $x = \mu$ on mõlemad tihedusfunktsioonid võrdsed, järelikult peab kehtima

$$\frac{\lambda_1}{\sqrt{2\pi\sigma_1^2}} = \frac{\lambda_2}{\sqrt{2\pi\sigma_2^2}}.$$

Integreeruvuse tingimusest järeldame, et kuna normaaljaotus on sümmeetriline, siis kumbki pool peab andma

$$\int_{-\infty}^{\mu} \lambda_1 f(x; \mu, \sigma_1) = \lambda_1 \frac{1}{2} \quad \text{ja} \quad \int_{\mu}^{\infty} \lambda_2 f(x; \mu, \sigma_2) = \lambda_2 \frac{1}{2},$$

seega

$$\frac{1}{2}\lambda_1 + \frac{1}{2}\lambda_2 = 1.$$

Eelnevat arvesse võttes saame, et

$$\lambda_1 + \lambda_2 = 2 \Leftrightarrow \lambda_1 + \lambda_1 \frac{\sigma_1}{\sigma_2} = 2 \Leftrightarrow \lambda_1 \left(1 + \frac{\sigma_2}{\sigma_1}\right) = 2 \Leftrightarrow \lambda_1 \frac{\sigma_1 + \sigma_2}{\sigma_1} = 2.$$

Analoogselt ka teise konstandiga toimetades leiame skaleerimiskonstandid

$$\lambda_1 = \frac{2\sigma_1}{\sigma_1 + \sigma_2} \quad \text{ja} \quad \lambda_2 = \frac{2\sigma_2}{\sigma_1 + \sigma_2}. \quad (9)$$

Seega saame kaheosalise normaaljaotuse tihedusfunktsiooni lõplikult defineerida järgnevalt:

$$g(x; \mu, \sigma_1, \sigma_2) = \begin{cases} \lambda_1 f(x; \mu, \sigma_1), & \text{kui } x \leq \mu, \\ \lambda_2 f(x; \mu, \sigma_2), & \text{kui } x > \mu. \end{cases} \quad (10)$$

Selleks, et andmete paiknemist paremini kirjeldada, on abi kvantiilidest. Mediaan ehk 0,5-kvantiil ($q_{0,5}$) on väärtus, millest mõlemale poole jääb sama palju andmeid. Näiteks sümmeetrilise normaaljaotuse mediaan on võrdne keskvaartusega, kuid kaheosalise normaaljaotuse puhul saame mediaani väärtusest teada, kummale poole on jaotus kaldu. Kvantiile saame leida kvantiilfunktsiooni kaudu, mis on omakorda pöördfunktsioon jaotusfunktsioonist. Seega paneme esmalt kirja kaheosalise normaaljaotuse jaotusfunktsiooni G . Selle tuletamiseks integreerime tihedusfunktsiooni g poollõigis $(-\infty, x]$. Kuna tihedusfunktsioon on defineeritud tükiti, jaguneb ka jaotusfunktsioon kaheks osaks vastavalt asukohaparametrile μ . Analoogselt tihedusfunktsioonile, peame jaotusfunktsiooni pooled skaleerima vastavalt kummagi kaheosalise jaotuse konstrueerimiseks kasutatud normaaljaotuse hajuvusele. Valemi (7) põhjal näeb kaheosalise jaotusfunktsiooni vasak pool välja järgmine:

$$G(x) = \int_{-\infty}^x \lambda_1 f(x; \mu, \sigma_1) dx = \lambda_1 \Phi \left(\frac{x - \mu}{\sigma_1} \right).$$

Punktis $x = \mu$ saame väärtuseks kaheosalise jaotuse vasakpoolse saba tõe-

näosusmassi. Kuna normaaljaotus on sümmeetriline, siis

$$G(\mu) = \lambda_1 \frac{1}{2} = \frac{2\sigma_1}{\sigma_1 + \sigma_2} \cdot \frac{1}{2} = \frac{\sigma_1}{\sigma_1 + \sigma_2}.$$

Jaotusfunktsiooni parem pool koosneb vasakpoolse saba tõenäosusmassist $G(\mu)$ ja sellele lisanduvast parema poole integraalist

$$G(x) = G(\mu) + \int_{\mu}^x \lambda_2 f(t; \mu, \sigma_2) dt.$$

Integraali $\int_{\mu}^x f(t; \mu, \sigma_2) dt$ saame kirjutada kujul $\Phi\left(\frac{x-\mu}{\sigma_2}\right) - \frac{1}{2}$ (lahutame $\frac{1}{2}$, sest muidu kasutaksime jaotuse vasakut poolt kaks korda). Kaheosalise jaotuse parem pool avaldub seega nii:

$$G(x) = \frac{\sigma_1}{\sigma_1 + \sigma_2} + \lambda_2 \left[\Phi\left(\frac{x-\mu}{\sigma_2}\right) - \frac{1}{2} \right].$$

Eelneva põhjal saame kaheosalise normaaljaotuse jaotusfunktsiooni kirja panna kujul:

$$G(x; \mu, \sigma_1, \sigma_2) = \begin{cases} \lambda_1 \Phi\left(\frac{x-\mu}{\sigma_1}\right), & \text{kui } x \leq \mu, \\ \frac{\sigma_1}{\sigma_1 + \sigma_2} + \lambda_2 \left[\Phi\left(\frac{x-\mu}{\sigma_2}\right) - \frac{1}{2} \right], & \text{kui } x > \mu. \end{cases} \quad (11)$$

Nüüd saame jaotusfunktsiooni pöörates kätte kaheosalise normaaljaotuse kvantiilfunktsiooni. Olgu q_p , kus $p \in [0, 1]$ kvantiil, mida otsime. Kriitiliseks punktiks on $G(\mu) = \frac{\sigma_1}{\sigma_1 + \sigma_2}$. Kui $p \leq \frac{\sigma_1}{\sigma_1 + \sigma_2}$, kasutame jaotusfunktsiooni esimest poolt

$$p = \lambda_1 \Phi\left(\frac{x-\mu}{\sigma_1}\right).$$

Jagame λ_1 -ga ja rakendame standardse normaaljaotuse kvantiilfunktsiooni

pöördfunktsiooni Φ^{-1} :

$$\Phi^{-1}\left(\frac{p}{\lambda_1}\right) = \frac{x - \mu}{\sigma_1}.$$

Siit leiame kvantiili q_p väärtuse:

$$x = \mu + \sigma_1 \Phi^{-1}\left(\frac{p}{\lambda_1}\right).$$

Juhul kui $p > \frac{\sigma_1}{\sigma_1 + \sigma_2}$, kasutame jaotusfunktsiooni paremat poolt

$$p = \frac{\sigma_1}{\sigma_1 + \sigma_2} + \lambda_2 \left[\Phi\left(\frac{x - \mu}{\sigma_2}\right) - \frac{1}{2} \right].$$

Teades, et $\frac{\sigma_1}{\sigma_1 + \sigma_2} = \frac{1}{2}\lambda_1$ ja $1 - \frac{1}{2}\lambda_1 = \frac{1}{2}\lambda_2$, saame vasakule jaotuse poolele analoogselt leida, et

$$x = \mu + \sigma_2 \Phi^{-1}\left(\frac{p - \frac{\sigma_1}{\sigma_1 + \sigma_2}}{\lambda_2} + \frac{1}{2}\right).$$

Kokku oleme leidnud kaheosalise normaaljaotuse kvantiilfunktsiooni

$$Q(p; \mu, \sigma_1, \sigma_2) = \begin{cases} \mu + \sigma_1 \Phi^{-1}\left(\frac{p}{\lambda_1}\right), & \text{kui } p \leq \frac{\sigma_1}{\sigma_1 + \sigma_2}, \\ \mu + \sigma_2 \Phi^{-1}\left(\frac{p - \frac{\sigma_1}{\sigma_1 + \sigma_2}}{\lambda_2} + \frac{1}{2}\right), & \text{kui } p > \frac{\sigma_1}{\sigma_1 + \sigma_2}. \end{cases} \quad (12)$$

2.3 Momentide leidmine

Selleks, et mõista seost kaheosalise normaaljaotusega juhusliku suuruse X keskväärtuse EX ja standarhälbe $\sqrt{DX}$ ning selle jaotuse moodustamisel aluseks olnud jaotuste parameetrite μ , σ_1 ja σ_2 vahel, leiame momendid. Need aitavad kummagi kaheosalise normaaljaotuse poole kuju paremini kirjeldada ning hiljem andmetega töötades täpseid parameetreid arvutada. Arvestame, et kaheosalise normaaljaotusega juhuslikul suurusel X , on parameetrist μ

vasakule poole jääv tõenäosusmass $\frac{1}{2}\lambda_1$ ning paremale poole jääv tõenäosusmass $\frac{1}{2}\lambda_2$. Märgime ära, et kui juhusliku suuruse keskväärtus võrdub nulliga, on tavalised momendid võrdsed tsentraalmomentidega. Lisaks absoluutsete tsentraalmomentide väärtus ei sõltu normaaljaotuse korral valemit (8) arvestades üldse keskväärtusest.

Leiame esmalt juhusliku suuruse X esimese momendi ehk keskväärtuse. Olgu meil juhuslik suurus Y selline, et $X = Y + \mu$. Seega teame, et $EX = E[Y + \mu] = EY + \mu$. Vaatleme sellist kaheosalise normaaljaotusega juhusliku suurust Y , mis koosneb normaaljaotusest keskväärtusega 0. Nii on meil lihtsam momente leida, arvestades, et juhul $\mu \neq 0$ erineb tulemus ainult parameetri μ nihke võrra. Seose (1) järgi ja kaheosalise normaaljaotuse tihefunktsioonist (10) saame, et

$$EY = \int_{-\infty}^{\infty} y \cdot g(y; 0, \sigma_1, \sigma_2) dy = \int_{-\infty}^0 y \lambda_1 f(y; 0, \sigma_1) dy + \int_0^{\infty} y \lambda_2 f(y; 0, \sigma_2) dy.$$

Normaaljaotuse sümmeetria põhjal teame, et

$$\int_{-\infty}^0 y f(y; 0, \sigma_1) dy = - \int_0^{\infty} y f(y; 0, \sigma_1) dy,$$

järelikult

$$EY = -\lambda_1 \int_0^{\infty} y f(y; 0, \sigma_1) dy + \lambda_2 \int_0^{\infty} y f(y; 0, \sigma_2) dy.$$

Märgime, et kui $Z \sim N(0, \sigma^2)$ siis valemist (8) saame, et

$$E[|Z - EZ|] = E|Z| = \int_{-\infty}^{\infty} |z| \cdot f(z; 0, \sigma) dz = \sigma \sqrt{\frac{2}{\pi}}.$$

Normaaljaotuse sümmeetria tõttu

$$\int_0^\infty z \cdot f(z; 0, \sigma) dz = \frac{1}{2} \sigma \sqrt{\frac{2}{\pi}},$$

seega

$$\begin{aligned} EY &= -\frac{1}{2} \lambda_1 \sigma_1 \sqrt{\frac{2}{\pi}} + \frac{1}{2} \lambda_2 \sigma_2 \sqrt{\frac{2}{\pi}} = -\frac{\sigma_1}{\sigma_1 + \sigma_2} \sigma_1 \sqrt{\frac{2}{\pi}} + \frac{\sigma_2}{\sigma_1 + \sigma_2} \sigma_2 \sqrt{\frac{2}{\pi}} \\ &= \sqrt{\frac{2}{\pi}} \left(\frac{\sigma_2^2 - \sigma_1^2}{\sigma_1 + \sigma_2} \right) = \sqrt{\frac{2}{\pi}} \left(\frac{(\sigma_2 - \sigma_1)(\sigma_2 + \sigma_1)}{\sigma_1 + \sigma_2} \right) = \sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1). \end{aligned}$$

Eelneva põhjal leiame, et kaheosalise normaaljaotusega juhusliku suuruse X keskväärtus võrdub:

$$EX = \sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) + \mu. \quad (13)$$

Juhul $\sigma_1 = \sigma_2$ on tegemist tavalise sümmeetrilise normaaljaotusega, kus $EX = \mu$. Samas näiteks juhul $\sigma_1 < \sigma_2$ nihkub keskväärtus EX suurusest μ paremale ehk suurema hajuvusega komponentide poole. Vastupidiselt, kui $\sigma_1 > \sigma_2$ nihkub keskväärtus vasakule.

Järgmiseks leiame juhusliku suuruse X dispersiooni. Kuna $DX = D[Y + \mu] = DY = E(Y)^2 - (EY)^2$, siis on meil analoogselt suurusele EY vaja leida ka $E(Y)^2$. Kuna normaaljaotus on sümmeetriline ning y^2 on paarisfunktsioon, siis

$$\begin{aligned} EY^2 &= \int_{-\infty}^\infty y^2 \cdot g(y; 0, \sigma_1, \sigma_2) dy = 2 \int_0^\infty y^2 \cdot g(y; 0, \sigma_1, \sigma_2) dy \\ &= \int_0^\infty y^2 \lambda_1 f(y; 0, \sigma_1) dy + \int_0^\infty y^2 \lambda_2 f(y; 0, \sigma_2) dy \\ &= \lambda_1 \int_0^\infty y^2 f(y; 0, \sigma_1) dy + \lambda_2 \int_0^\infty y^2 f(y; 0, \sigma_2) dy. \end{aligned}$$

Integraal kujul $\int_0^\infty y^2 f(y; 0, \sigma_1) dy$ annab seose (1) järgi poole normaaljaotu-

se $N(0, \sigma_1^2)$ teisest momendist ehk antud juhul dispersioonist. Analoogselt $\int_0^\infty y^2 f(y; 0, \sigma_2) dy$ annab $\frac{1}{2}$ normaalkaotuse $N(0, \sigma_2^2)$ dispersioonist, seega

$$E(Y)^2 = \lambda_1 \frac{1}{2} \sigma_1^2 + \lambda_2 \frac{1}{2} \sigma_2^2 = \frac{\sigma_1^3 + \sigma_2^3}{\sigma_1 + \sigma_2}.$$

Järelikult kaheosalise normaalkaotusega juhusliku suuruse X dispersioon võrdub:

$$DX = DY = \frac{\sigma_1^3 + \sigma_2^3}{\sigma_1 + \sigma_2} - \left(\sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) \right)^2 = \frac{\sigma_1^3 + \sigma_2^3}{\sigma_1 + \sigma_2} - \frac{2}{\pi} (\sigma_2 - \sigma_1)^2. \quad (14)$$

Asümmeetriakordaja jaoks peame leidma juhusliku suuruse X kolmanda tsentraalmomendi. Märgime, et tsentraalmoment ei sõltu nihke μ väärtusest, sest $X - EX = (Y + \mu) - (E(Y + \mu)) = Y + \mu - EY - \mu = Y - EY$. Järelikult valemi (2) põhjal $E[(X - EX)^3] = E[(Y - EY)^3] = E(Y)^3 - (EY)^3 - 3EYDY$, seega meil on lisaks olemasolevale vaja leida veel juhusliku suuruse Y kolmas moment $E(Y)^3$, mida saab teha analoogselt esimese momendi leidmisele:

$$\begin{aligned} E(Y)^3 &= \int_{-\infty}^{\infty} y^3 \cdot g(y; 0, \sigma_1, \sigma_2) dy \\ &= -\lambda_1 \int_0^{\infty} y^3 f(y; 0, \sigma_1) dy + \lambda_2 \int_0^{\infty} y^3 f(y; 0, \sigma_2) dy. \end{aligned}$$

Märgime, et kui $Z \sim N(0, \sigma^2)$ siis valemist (8) saame, et

$$E[|Z - EZ|^3] = E|Z|^3 = \int_{-\infty}^{\infty} |z|^3 f(z; 0, \sigma) dz = 2\sigma^3 \sqrt{\frac{2}{\pi}}.$$

Kuna integraal vahemikus $[0, \infty)$ on täpselt pool absoluutsest tsentraalmomendist, siis

$$\int_0^{\infty} z^3 f(z; 0, \sigma) dz = \frac{1}{2} \left(2\sigma^3 \sqrt{\frac{2}{\pi}} \right) = \sigma^3 \sqrt{\frac{2}{\pi}},$$

seega

$$\begin{aligned} E(Y)^3 &= -\lambda_1 \sigma_1^3 \sqrt{\frac{2}{\pi}} + \lambda_2 \sigma_2^3 \sqrt{\frac{2}{\pi}} = \sqrt{\frac{2}{\pi}} \left(-\frac{2\sigma_1}{\sigma_1 + \sigma_2} \sigma_1^3 + \frac{2\sigma_2}{\sigma_1 + \sigma_2} \sigma_2^3 \right) \\ &= \sqrt{\frac{2}{\pi}} \left(\frac{-2\sigma_1^4 + 2\sigma_2^4}{\sigma_1 + \sigma_2} \right) = 2\sqrt{\frac{2}{\pi}} \left(\frac{\sigma_2^4 - \sigma_1^4}{\sigma_1 + \sigma_2} \right). \end{aligned}$$

Nüüd saame leida kolmanda tsentraalmomendi:

$$\begin{aligned} E[(X - EX)^3] &= 2\sqrt{\frac{2}{\pi}} \left(\frac{\sigma_2^4 - \sigma_1^4}{\sigma_1 + \sigma_2} \right) - \left(\sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) \right)^3 \\ &\quad - 3\sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) \left(\frac{\sigma_1^3 + \sigma_2^3}{\sigma_1 + \sigma_2} - \frac{2}{\pi} (\sigma_2 - \sigma_1)^2 \right) \\ &= 2\sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) (\sigma_1^2 + \sigma_2^2) - \frac{2}{\pi} \sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1)^3 \\ &\quad - 3\sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) \left(\sigma_1^2 - \sigma_1 \sigma_2 + \sigma_2^2 - \frac{2}{\pi} (\sigma_2 - \sigma_1)^2 \right) \\ &= \sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) \left[2(\sigma_1^2 + \sigma_2^2) - \frac{2}{\pi} (\sigma_2 - \sigma_1)^2 - 3(\sigma_1^2 - \sigma_1 \sigma_2 + \sigma_2^2) + \frac{6}{\pi} (\sigma_2 - \sigma_1)^2 \right] \\ &= \sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) \left[2\sigma_1^2 + 2\sigma_2^2 - 3\sigma_1^2 + 3\sigma_1 \sigma_2 - 3\sigma_2^2 + \frac{4}{\pi} (\sigma_2 - \sigma_1)^2 \right] \\ &= \sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) \left[-\sigma_1^2 + 3\sigma_1 \sigma_2 - \sigma_2^2 + \frac{4}{\pi} (\sigma_1^2 - 2\sigma_1 \sigma_2 + \sigma_2^2) \right] \\ &= \sqrt{\frac{2}{\pi}} (\sigma_2 - \sigma_1) \left[\left(\frac{4}{\pi} - 1 \right) (\sigma_1^2 + \sigma_2^2) + \left(3 - \frac{8}{\pi} \right) \sigma_1 \sigma_2 \right]. \end{aligned}$$

Järsakuskordaja γ_2 jaoks peame leidma neljandat järku tsentraalmomendi $E[(X - EX)^4]$. Ka järsakuskordaja ei sõltu asukohaparametrist μ , seega piisab kui leiame juhusliku suuruse Y neljanda tsentraalmomendi valemi (4) põhjal:

$$E(Y)^4 = \int_{-\infty}^{\infty} y^4 g(y; 0, \sigma_1, \sigma_2) dy = \lambda_1 \int_0^{\infty} y^4 f(y; 0, \sigma_1) dy + \lambda_2 \int_0^{\infty} y^4 f(y; 0, \sigma_2) dy.$$

Märgime, et kui $Z \sim N(0, \sigma^2)$ siis valemist (8) saame, et

$$E[|Z - EZ|^4] = E|Z|^4 = \int_{-\infty}^{\infty} |z|^4 f(z; 0, \sigma) dz = 3\sigma^4.$$

Integraal vahemikus $[0, \infty)$ on pool absoluutsest tsentraalmomendist

$$\int_0^{\infty} y^4 f(y; 0, \sigma) dy = \frac{3}{2}\sigma^4.$$

Järelikult:

$$\begin{aligned} E(Y)^4 &= \frac{2\sigma_1}{\sigma_1 + \sigma_2} \cdot \frac{3}{2}\sigma_1^4 + \frac{2\sigma_2}{\sigma_1 + \sigma_2} \cdot \frac{3}{2}\sigma_2^4 = \frac{3\sigma_1^5 + 3\sigma_2^5}{\sigma_1 + \sigma_2} = 3 \left(\frac{\sigma_1^5 + \sigma_2^5}{\sigma_1 + \sigma_2} \right) \\ &= \frac{3(\sigma_1 + \sigma_2)(\sigma_1^4 - \sigma_1^3\sigma_2 + \sigma_1^2\sigma_2^2 - \sigma_1\sigma_2^3 + \sigma_2^4)}{\sigma_1 + \sigma_2} \\ &= 3(\sigma_1^4 - \sigma_1^3\sigma_2 + \sigma_1^2\sigma_2^2 - \sigma_1\sigma_2^3 + \sigma_2^4). \end{aligned}$$

Eelneva põhjal tuletame juhusliku suuruse X neljandat järku tsentraalmomendi:

$$\begin{aligned} E[(X - EX)^4] &= \underbrace{3(\sigma_1^4 - \sigma_1^3\sigma_2 + \sigma_1^2\sigma_2^2 - \sigma_1\sigma_2^3 + \sigma_2^4)}_{E(Y)^4} \\ &\quad - 4 \cdot \underbrace{\left[2\sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1)(\sigma_1^2 + \sigma_2^2) \right]}_{E(Y)^3} \cdot \underbrace{\left[\sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1) \right]}_{EY} \\ &\quad + 6 \cdot \underbrace{(\sigma_1^2 - \sigma_1\sigma_2 + \sigma_2^2)}_{E(Y)^2} \cdot \underbrace{\left[\sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1) \right]^2}_{(EY)^2} - 3 \cdot \underbrace{\left[\sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1) \right]^4}_{(EY)^4} \\ &= 3(\sigma_1^4 - \sigma_1^3\sigma_2 + \sigma_1^2\sigma_2^2 - \sigma_1\sigma_2^3 + \sigma_2^4) - \frac{16}{\pi}(\sigma_2 - \sigma_1)^2(\sigma_1^2 + \sigma_2^2) \\ &\quad + \frac{12}{\pi}(\sigma_1^2 - \sigma_1\sigma_2 + \sigma_2^2)(\sigma_2 - \sigma_1)^2 - \frac{12}{\pi^2}(\sigma_2 - \sigma_1)^4. \end{aligned}$$

2.4 Asümmeetria- ning järsakuskordaja

Leiame juhusliku suuruse X jaotuse asümmeetriakordaja asendades momendid valemisse (3):

$$\gamma_1 = \frac{\sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1) \left[\left(\frac{4}{\pi} - 1 \right) (\sigma_1^2 + \sigma_2^2) + \left(3 - \frac{8}{\pi} \right) \sigma_1 \sigma_2 \right]}{\left[\sigma_1^2 - \sigma_1 \sigma_2 + \sigma_2^2 - \frac{2}{\pi}(\sigma_2 - \sigma_1)^2 \right]^{3/2}}. \quad (15)$$

Juhul kui $\sigma_1 = \sigma_2$, siis asümmeetriakordaja $\gamma_1 = 0$, mis on loogiline, sest tegemist oleks tavalise sümmeetrilise normaaljaotusega. Kui $\sigma_2 > \sigma_1$ on tegemist positiivse asümmeetriaga (Joonis 2), sest kaheosalise jaotuse parempoolne saba on pikem. Analoogselt, kui $\sigma_1 > \sigma_2$ on tegemist negatiivse asümmeetriaga.

Andmetega sobitamisel peab ka jälgima, et arvutatud asümmeetriakordaja väärtus jääks lubatud piiridesse, vastasel juhul ei saa kaheosalist normaaljaotust nende andmete peal kasutada. Asümmeetriakordaja piiride leidmiseks eeldame, et kaheosaline normaaljaotus on maksimaalselt parempoolse asümmeetriaga ehk vasaku poole hajuvus läheneb nullile:

$$\lim_{\sigma_1 \rightarrow 0} \gamma_1 = \frac{\sqrt{\frac{2}{\pi}} \sigma_2 \left(\frac{4}{\pi} - 1 \right) \sigma_2^2}{(\sigma_2^2 - \frac{2}{\pi} \sigma_2^2)^{3/2}} = \frac{\sigma_2^3 \sqrt{\frac{2}{\pi}} \left(\frac{4}{\pi} - 1 \right)}{\sigma_2^{6/2} \left(1 - \frac{2}{\pi} \right)^{3/2}} = \frac{\sqrt{\frac{2}{\pi}} \left(\frac{4}{\pi} - 1 \right)}{\left(1 - \frac{2}{\pi} \right)^{3/2}} \approx 0,995. \quad (16)$$

Kui jaotus oleks maksimaalselt vasakpoolse asümmeetriaga ehk $\sigma_2 \rightarrow 0$, saaksime analoogse lahendusega asümmeetriakordaja väärtuseks $\gamma_1 \approx -0,995$. Tulemused vastavad pool-normaaljaotuse (ingl *half-normal distribution*) asümmeetriakordajale, mis on ka loogiline, sest kui ühe poole hajuvus läheneb nullile, jääb alles pool normaaljaotust. Seega oleme leidnud kaheosalise normaaljaotuse asümmeetriakordaja piirid $\gamma_1 \in (-0,995; 0,995)$.

Leiame ka juhusliku suuruse X jaotuse järsakuskordaja asendades momendid

valemissse (5):

$$\gamma_2 = \frac{3(\sigma_1^4 - \sigma_1^3\sigma_2 + \sigma_1^2\sigma_2^2 - \sigma_1\sigma_2^3 + \sigma_2^4) - \frac{16}{\pi}(\sigma_2 - \sigma_1)^2(\sigma_1^2 + \sigma_2^2) + \frac{12}{\pi}(\sigma_1^2 - \sigma_1\sigma_2 + \sigma_2^2)(\sigma_2 - \sigma_1)^2 - \frac{12}{\pi^2}(\sigma_2 - \sigma_1)^4}{[\sigma_1^2 - \sigma_1\sigma_2 + \sigma_2^2 - \frac{2}{\pi}(\sigma_2 - \sigma_1)^2]^2}. \quad (17)$$

Juhul $\sigma_1 = \sigma_2$ saame, et $\gamma_2 = 3$, mis võrdub tavalise sümmeetrilise normaaljaotuse järsakuskordajaga (ingl *mesokurtic distribution*). Kui $\gamma_2 < 3$ on jaotus normaaljaotusest lamedama tipuga (ingl *platykurtic distribution*) ehk kergemate sabadega ning kui $\gamma_2 > 3$ on jaotus normaaljaotusest teravama tipuga (ingl *leptokurtic distribution*) ehk raskemate sabadega. (Balanda ja MacGillivray, 1988)

Kombineerides kaks normaaljaotust üheks tuleb saadud kaheosalise jaotuse tipp teravam normaaljaotuse tipust, seega järsakuskordaja alumiseks piiriks on $\gamma_2 = 3$. Ülemise piiri leidmiseks vaatleme taas olukorda, kus asümmeetria on maksimaalne:

$$\lim_{\sigma_1 \rightarrow 0} \gamma_2 = \frac{3\sigma_2^4 - \frac{16}{\pi}(\sigma_2^2)^2 + \frac{12}{\pi}\sigma_2^2(\sigma_2^2) - \frac{12}{\pi^2}(\sigma_2^2)^2}{(\sigma_2^2 - \frac{2}{\pi}\sigma_2^2)^2} = \frac{\sigma_2^4(3 - \frac{4}{\pi} - \frac{12}{\pi^2})}{\sigma_2^4(1 - \frac{2}{\pi})^2} \approx 3,869.$$

Vastus tuleb sama kui võtta $\sigma_2 \rightarrow 0$. Ka seekord võrdub saadud tulemus pool-normaaljaotuse järsakuskordajaga.

Tasub ka vaadelda juhtu, kus üks hajuvustest nt σ_1 on fikseeritud ning teine läheneb lõpmatusele $\sigma_2 \rightarrow \infty$. Defineerime suhtarvu $k = \frac{\sigma_1}{\sigma_2}$. Jagades järsakuskordaja valemis (17) lugeja ja nimetaja läbi liikmega σ_2^4 , saame avaldise:

$$\gamma_2 = \frac{3(k^4 - k^3 + k^2 - k + 1) - \frac{16}{\pi}(1 - k)^2(k^2 + 1) + \frac{12}{\pi}(k^2 - k + 1)(1 - k)^2 - \frac{12}{\pi^2}(1 - k)^4}{[k^2 - k + 1 - \frac{2}{\pi}(1 - k)^2]^2}. \quad (18)$$

Kuna $k \rightarrow 0$, siis:

$$\lim_{\sigma_2 \rightarrow \infty} \gamma_2 = \frac{3 - \frac{16}{\pi} + \frac{12}{\pi} - \frac{12}{\pi^2}}{\left[1 - \frac{2}{\pi}\right]^2} = \frac{3 - \frac{4}{\pi} - \frac{12}{\pi^2}}{\left(1 - \frac{2}{\pi}\right)^2} \approx 3,869. \quad (19)$$

Järelikult mõjutab järsakuskordajat mitte hajuvuste suuruse vahe, vaid nende omavaheline suhe $\frac{\sigma_1}{\sigma_2}$. Oleme leidnud järsakuskordaja piirid kaheosalise normaalkaotuse jaoks $\gamma_2 \in [3; 3,869)$.

Nii asümmeetria- kui ka järsakuskordaja piirid kaheosalise normaalkaotuse jaoks on suhteliselt kitsad. Sellest saame järeldada, et antud meetod sobib kasutamiseks vaid väikese asümmeetriaga ning kergelt teravam tipuga andmetel.

2.5 Numbriline näide

Näitlikustamiseks eespool leitud tulemusi, rakendame konkreetseid parameetreid, et illustreerida kaheosalise normaalkaotuse kuju. Olgu meil antud kaks normaalkaotust $N \sim (\mu = 2, \sigma_1^2 = 0,8^2)$ ja $N \sim (\mu = 2, \sigma_2^2 = 0,4^2)$. Koostame nende põhjal juhusliku suuruse X kaheosalise normaalkaotuse.

Esmalt leiame valemi (9) põhjal skaleerimiskonstandid:

$$\lambda_1 = \frac{2 \cdot 0,8}{0,8 + 0,4} = \frac{1,6}{1,2} = \frac{4}{3}, \quad \lambda_2 = \frac{2 \cdot 0,4}{0,8 + 0,4} = \frac{0,8}{1,2} = \frac{2}{3}.$$

Kuna $\lambda_1 > \lambda_2$, siis vasakpoolne osa on kõrgemale skaleeritud, et tagada pidevus jaotuse tipus.

Järgmiseks leiame valemi (13) põhjal kaheosalise normaalkaotusega juhusliku suuruse X keskväärtuse ja valemi (14) põhjal dispersiooni:

$$EX = 2 + \sqrt{\frac{2}{\pi}}(0,4 - 0,8) = 2 - 0,4 \cdot \sqrt{\frac{2}{\pi}} \approx 2 - 0,3192 = 1,6808,$$

$$DX = \frac{0,8^3 + 0,4^3}{0,8 + 0,4} - \frac{2}{\pi}(0,4 - 0,8)^2 = \frac{0,512 + 0,064}{1,2} - \frac{2}{\pi}(0,16) \approx 0,3781.$$

Nagu näha, nihutab suurema hajuvusega vasakpoolne saba keskvärtuse jaotuse tipust ($\mu = 2$) vasakule. Dispersioon jääb umbes komponentide dispersioonide $\sigma_1^2 = 0,64$ ja $\sigma_2^2 = 0,16$ keskele.

Kuna $\sigma_1 > \sigma_2$, ootame asümmeetriakordajast negatiivset tulemust. Valemi (15) põhjal:

$$\gamma_1 = \frac{\sqrt{\frac{2}{\pi}}(-0,4) \left[\left(\frac{4}{\pi} - 1\right)(0,8^2 + 0,4^2) + \left(3 - \frac{8}{\pi}\right)(0,8 \cdot 0,4) \right]}{(0,3781)^{3/2}} \approx -0,562.$$

Tulemus $\gamma_1 \approx -0,562$ kinnitab mõõdukalt vasakpoolset asümmeetriat. Asendades parameetrid valemisse (17), saame kaheosalise normaaljaotusega juhusliku suruse X järsakuskordajaks $\gamma_2 \approx 3,21$. See väärtus on suurem kui 3, mis tähendab, et jaotuse tipp on teravam kui vastaval sümmeetrilisel normaaljaotusel.

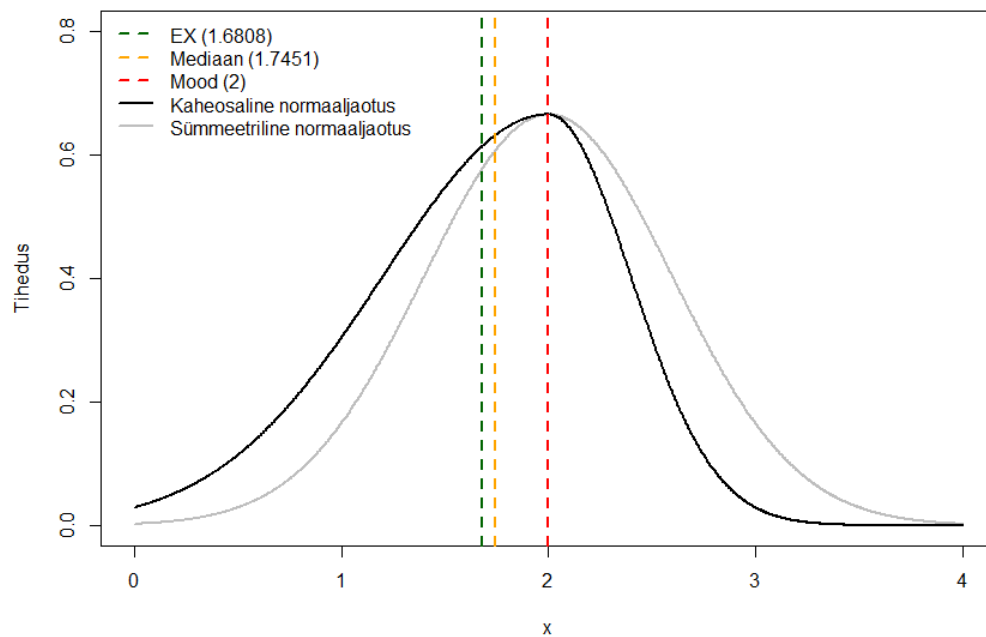
Leiame kaheosalise jaotuse mediaani $q_{0,5}$. Näeme, et jaotuse kriitiliseks punktiks tuleb:

$$G(2) = \frac{\sigma_1}{\sigma_1 + \sigma_2} = \frac{0,8}{1,2} = \frac{2}{3}.$$

Kuna otsitav kvantiil $p = 0,5$ on väiksem kui $\frac{2}{3} = 0,667$, asub mediaan jaotuse vasakpoolses osas, seega kasutame kvantiilfunktsiooni (12) vastavalt:

$$q_{0,5} = 2 + 0,8 \cdot \Phi^{-1}\left(\frac{3 \cdot 0,5}{4}\right) = 2 + 0,8 \cdot \Phi^{-1}(0,375).$$

Standardse normaaljaotuse tabelist leiame $\Phi^{-1}(0,375) \approx -0,3186$, seega mediaani väärtus tuleb: $q_{0,5} = 2 + 0,8 \cdot (-0,3186) \approx 2 - 0,2549 = 1,7451$.



Joonis 3: Kaheosaline normaaljaotus, $EX = 1,6808$ ja $DX = 0,3781$

Nagu näha ka joonisel 3, kehtib antud näites negatiivsele asümmeetriaale oma-
ne järjestus:

$$EX(1,6808) < \text{mediaan}(1,7451) < \text{mood}(2).$$

Märgime, et sümmeetrilise normaaljaotuse korral oleksid kõik need näitajad
omavahel võrdsed. See illustreerib, kuidas kaheosalise normaaljaotuse para-
meetrid σ_1 ja σ_2 määravad jaotuse kuju.

3 Jaotuse sobitamine andmetele

Selles peatükis rakendame eelnevalt saadud teoreetilisi tulemusi praktikas. Esmalt anname ülevaate parameetrite hindamisest. Seejärel sobitame kaheosalist normaaljaotust nii genereeritud kui ka päris andmetele ning võrdleme mõlemal juhul tulemusi teiste tuntud jaotustega.

3.1 Parameetrite hindamine

Selleks, et kaheosalist normaaljaotust andmetele sobitada, peame valimi põhjal leidma jaotuse koostamiseks vajalike parameetrite hinnangud. Vaatleme sõltumatute sama jaotusega juhuslike suuruste valimit $(X_1, \dots, X_n)$, mille realisatsioon on $\mathbf{x} = (x_1, \dots, x_n)$. Kasutame suurima tõepära meetodit, mille eesmärgiks on leida parameetrite μ , σ_1 ja σ_2 väärtused nii, et tõepärafunktsioon $L(\mu, \sigma_1, \sigma_2 | \mathbf{x})$ oleks maksimaalne ehk olemasolevate andmete $\mathbf{x}$ esinemine oleks kõige tõenäolisem. Praktikas on lihtsam kasutada tõepärafunktsiooni logaritmi:

$$\ell(\mu, \sigma_1, \sigma_2 | \mathbf{x}) = \sum_{i=1}^n \ln g(x_i; \mu, \sigma_1, \sigma_2),$$

kus kaheosalise normaaljaotuse tihedusfunktsioon g on defineeritud valemiga (10) ja n on valimimaht.

Tõepärafunktsiooni logaritmi maksimeerimisel üle võimalike parameetrite μ , σ_1 ja σ_2 väärtuste, saame nende suurima tõepära hinnangud:

$$(\hat{\mu}, \hat{\sigma}_1, \hat{\sigma}_2)_{MLE} = \arg \max_{\mu, \sigma_1, \sigma_2} \ell(\mu, \sigma_1, \sigma_2 | \mathbf{x}).$$

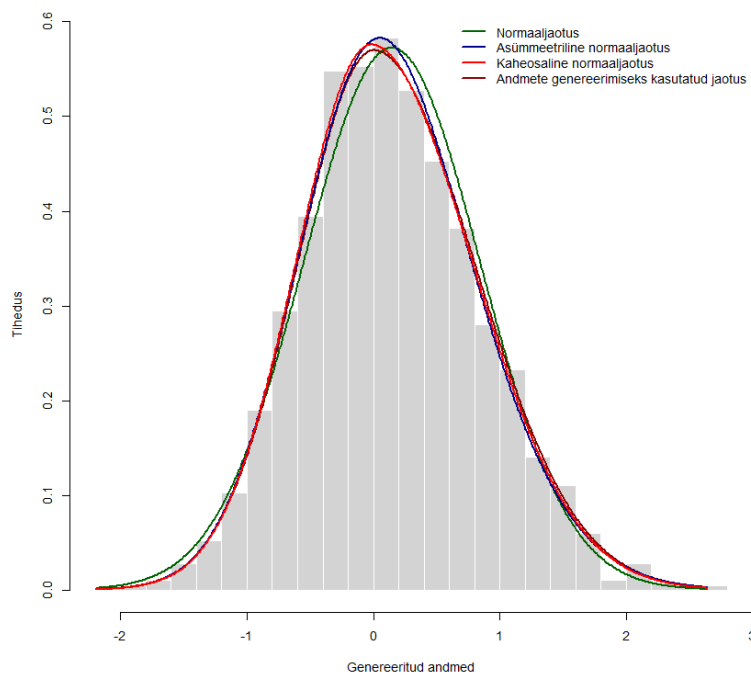
Hinnangute väljaarvutamiseks kasutame rakendustarkvara R sisseehitatud paketi *stats4* funktsiooni *mle* (R Core Team, 2025).

3.2 Jaotuse sobitamine genereeritud andmetele

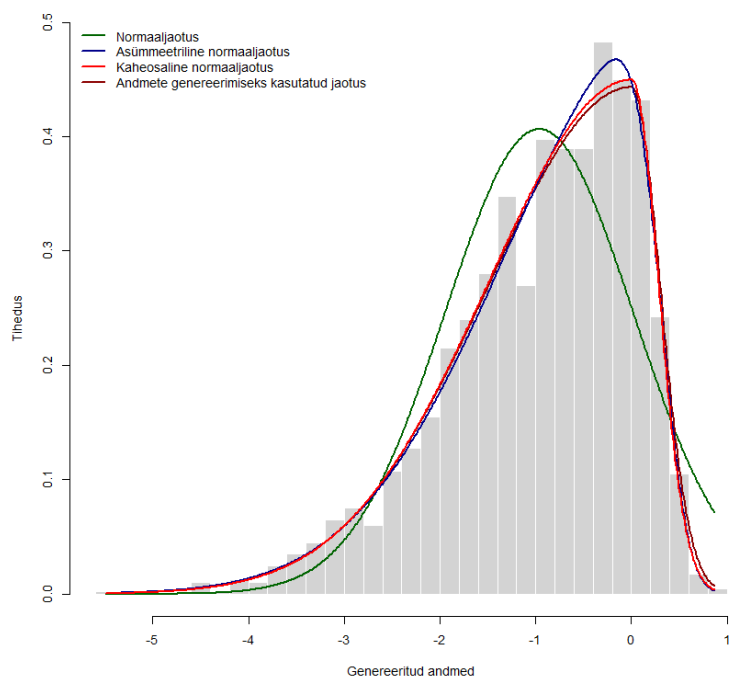
Esmalt proovime läbi kaheosalise normaaljaotuse parameetrite hindamist genereeritud andmetel. See tähendab, et teame andmete õiget jaotust ning saame parameetrite hinnangute täpsust kontrollida. Andmeid genereerime kahest erineva hajuvusega pool-normaaljaotusest, paigutades neid vastavalt kummalegi poole moodi. Nii saamegi kaheosalise normaaljaotusega andmed. Jaotuse sobivust andmetele hindame sobivusnäitajaga AIC; mida madalam on AIC, seda paremini sobib jaotus andmetele. Näidiskood andmete genereerimisest, parameetrite hindamisest ja AIC-i arvutamisest on leitav lisast [A](#).

Genereeritud andmetele sobitame kolm jaotust: kaheosaline normaaljaotus, normaaljaotus ning asümmeetriline normaaljaotus (*skew-normal distribution*). Viimane on Adelchi Azzalini poolt tuletatud jaotus, mis võtab tavalise normaaljaotuse ning muudab asümmeetriaparametri α abil ühe saba raskemaks (Azzalini, 1985). Kaheosalist ja asümmeetrilist normaaljaotust on hea omavahel võrrelda, sest mõlemad rõhuvad asümmeetrilistele andmetele, säilitades normaaljaotusele omase kellukese kuju. Asümmeetrilise jaotuse tihedusfunktsiooni jaoks kasutame eraldi rakendustarkvara R paketti *sn* (Azzalini, 2026).

Joonisel 4 olevate andmete genereerimiseks kasutame parameetreid $\mu = 0$, $\sigma_1 = 0,6$ ja $\sigma_2 = 0,8$. Nende andmete asümmeetriakordaja tuleb $\gamma_1 = 0,225$ ja järsakuskordaja $\gamma_2 = 3,036$. Parameetrite hindamisel proovime algväärtusena mitmeid parameetrite komplekte. Saame kaheosalise normaaljaotuse parameetrite hinnanguteks alati $\hat{\mu} = -0,022$, $\hat{\sigma}_1 = 0,588$ ja $\hat{\sigma}_2 = 0,798$. Edaspidi kasutame algväärtusena andmete tiheduse tuumahinnangu maksimumpunkti kaudu leitud moodi ning sellest kummalgi poole jäänud standardhälbeid, sest tavaliselt ei ole andmete jaotus meile teada.



Joonis 4: Kaheosalise normaaljaotusega andmed: $\mu = 0$, $\sigma_1 = 0,6$, $\sigma_2 = 0,8$



Joonis 5: Kaheosalise normaaljaotusega andmed: $\mu = 0$, $\sigma_1 = 1,5$, $\sigma_2 = 0,3$

Tabel 1: Genereeritud andmetele sobitatud jaotuste AIC

	Kaheosaline-	Asümmeetriline-	Normaaljaotus
Joonis 4	4 214,17	4 216,14	4 233,41
Joonis 5	5 202,84	5 205,08	5 603,79

Joonisel 5 olevate andmete genereerimiseks kasutame parameetreid $\mu = 0$, $\sigma_1 = 1,5$ ja $\sigma_2 = 0,3$. Nende andmete asümmeetriakordaja tuleb $\gamma_1 = -0,841$ ja järsakuskordaja $\gamma_2 = 3,562$. Kaheosalise normaaljaotuse parameetrite hinnangud tulevad seekord $\hat{\mu} = 0,002$, $\hat{\sigma}_1 = 1,496$ ja $\hat{\sigma}_2 = 0,279$. Võime öelda, et parameetrite hindamine suurime tõepära meetodil on õnnestunud.

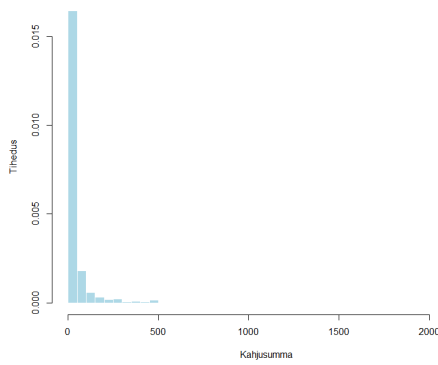
Tabelist 1 näeme, et genereeritud andmetele sobib kaheosaline normaaljaotus kõige paremini, mis on ka oodatav. Asümmeetrilise normaaljaotuse AIC on mõlemal näite korral umbes kahe punkti võrra kõrgem ehk sobitub natuke vähem. Normaaljaotuse AIC on palju suurem, mis on ka loogiline, sest see ei kajasta andmete asümmeetriat.

3.3 Jaotuse sobitamine reaalsele kahjuandmetele

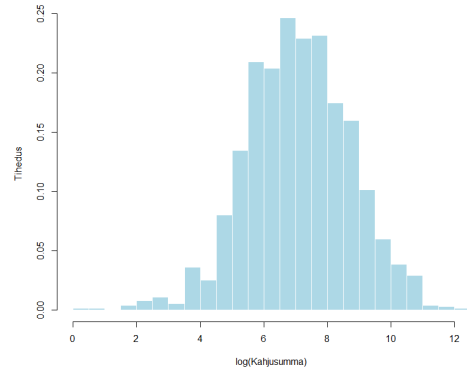
Järgnevas alapeatükis katsetame kaheosalist normaaljaotust päris andmetel. Tehisintellekti abiga leidsime kümmekond andmestikku, mida võiks kaheosaline normaaljaotus hästi kirjeldada. Sobitasime R-is erinevaid jaotusi nendele andmetele ning kõigil juhtudel oli kas asümmeetrilise normaaljaotuse või asümmeetrilise t-jaotuse AIC madalam. Paraku ei õnnestunud leida andmestikku, millele kaheosaline normaaljaotus sobiks kõige paremini. Seetõttu toome tulemused välja vaid ühe reaalse andmestiku kohta.

Martin Eling on 2012. aastal kirjutanud artikli sellest, kui hästi sobituvad asümmeetrilised jaotused kahjuandmetele (Eling, 2012). Ta kasutas rakendustarkvara R sisseehitatud paketi *copula* andmestikku nimega *loss* (Ho-

fert *et al.*, 2026), mis sisaldab USA kahjusummasid (USA dollarites) aastast 1998. Andmestikus on 1 500 kahjusummat, millest minimaalne kahju on 10, maksimaalne on 2 173 595, keskmine kahju suurus on 41 208 ning mediaan on 12 000 USA dollarit. Kaheosaline normaaljaotus transformeerimata kujul andmetele (Joonis 6) ei sobi, sest artiklis esitatud asümmeetriakordaja $\gamma_1 = 9,15$ on väljaspool kaheosalise normaaljaotuse asümmeetriakordaja piire $(-0,995; 0,995)$. Eling viis kahjuandmetel läbi logaritmilise teisenduse, mille tulemusena $\gamma_1 = -0,15$, seega logaritmitud andmetele (Joonis 7) võime kaheosalist normaaljaotust sobitada.

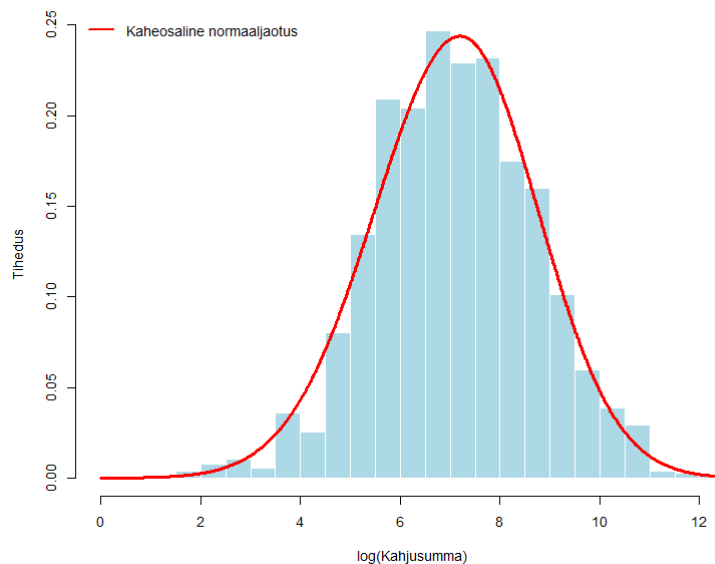


Joonis 6: Kahjuandmed



Joonis 7: Logaritmitud kahjuandmed

Esmalt peame hindama kaheosalise normaaljaotuse parameetrid kahjuandmete põhjal. Saame, et $\hat{\mu} = 7,210$, $\hat{\sigma}_1 = 1,724$ ja $\hat{\sigma}_2 = 1,549$. Parameetrid viitavad jaotuse kergele vasakpoolsele asümmeetriale, mida võib aimata ka jooniselt 7. Jooniselt 8 näeme, et kaheosalise normaaljaotuse tihedusfunktsioon katab andmeid päris hästi. Arvutades $\hat{\sigma}_1$ ja $\hat{\sigma}_2$ põhjal välja asümmeetria- ja järsakuskordaja, saame tulemuseks vastavalt $\gamma_1 = -0,085$ ja $\gamma_2 = 3,005$. Kordajate väärtused kajastavad kaheosalise normaaljaotuse tihedusfunktsiooni kuju joonisel 8. Järelikult oli parameetrite hindamine edukas ning kaheosalist normaaljaotust saaks kasutada nende andmete jaotuse kirjeldamiseks.



Joonis 8: Andmetele sobitatud kaheosaline normaaljaotus

Eling on USA kahjuandmetele sobitanud asümmeetrilist normaaljaotust ning asümmeetrilist t -jaotust, võrreldes tulemusi teiste jaotuste sobivusnäitajaga AIC. Artiklis uuritud jaotuste ja kaheosalise normaaljaotuse AIC suurused on välja toodud lisas B. Tulemuste põhjal näeme, et kaheosalist normaaljaotust küll saab kasutada, kuid see ei ole andmetega kõige paremini sobituv jaotus. Kaheosalise normaaljaotuse sobivusnäitaja AIC on 5 740,38, mis paigutab selle jaotuste paremusjärjestuses kümnendale kohale 20-st, tavalise normaaljaotuse ette. Kõige täpsemalt kirjeldaks USA kahjuandmeid lisas B esitatud jaotustest asümmeetriline normaaljaotus.

Kokkuvõte

Käesoleva bakalaureusetöö eesmärgiks oli uurida kaheosaliste jaotuste meetodit, keskendudes kaheosalisele normaaljaotusele. Töö käigus konstrueeriti kaheosalise normaaljaotuse tihedus-, jaotus-, ja kvantiilfunktsioon; tuletati momendid ning asümmeetria- ja järsakuskordaja. Samuti testiti jaotuse sobivust nii genereeritud kui ka päris andmetel.

Teoreetilisest osast selgus, et kaheosalise normaaljaotuse kuju on määratud tema koostamiseks kasutatud normaaljaotuste hajuvuste põhjal. Kui vasakul pool kasutatud normaaljaotuse hajuvus on suurem, siis on ka kaheosaline normaaljaotus vasakpoolse asümmeetriaga. Lisaks leiti, et kaheosalise normaaljaotuse asümmeetriakordaja peab jääma vahemikku $(-0,995; 0,995)$ ning järsakuskordaja poollõiku $[3; 3,869)$. See tähendab, et kaheosaline normaaljaotus sobib kasutamiseks andmetel, mis on kerge asümmeetriaga ning natukene normaaljaotusest teravama tipuga (*leptokurtic distribution*).

Praktilises osas tutvustati kaheosalise normaaljaotuse parameetrite hindamist ning sobitati jaotust andmetele. Genereeritud andmete põhjal kontrolliti parameetrite hindamise täpsust, mis osutus edukaks. Kaheosaline jaotus sobis AIC näitaja põhjal nendele andmetele asümmeetrilisest normaaljaotusest ning tavalisest normaaljaotusest paremini, mis oli ka oodatav. Lisaks sobitati kaheosalist normaaljaotust reaalsele kahjuandmetele ja võrreldi sobivust 19 teise jaotusega. Kaheosaline normaaljaotus jäi paremusjärjestuses keskele: seda saaks küll andmete kirjeldamiseks kasutada, kuid näiteks asümmeetriline normaaljaotus sobis konkreetsetele andmetele paremini.

Töös tutvustati jaotust, mida saab kasutada asümmeetriliste andmete kirjeldamiseks. Tehisintellekti abiga leiti kümme andmestikku, mis oleksid potentsiaalselt võinud olla kaheosalise normaaljaotusega. Paraku selgus kõi-

gil juhtudel, et asümmeetriline normaaljaotus või asümmeetriline t-jaotus sobivad andmetele paremini. Edasiarendusena tasuks konstrueerida ka kaheosaline t-jaotus, sest paljudel andmestikel on kaheosalise normaaljaotuse jaoks liiga rasked sabad.

Kasutatud allikad

- Azzalini, A. (1985). “A Class of Distributions Which Includes the Normal Ones”. *Scandinavian Journal of Statistics* 12.2, lk. 171–178. URL: <https://www.jstor.org/stable/4615982>.
- Azzalini, A. (2026). *The R package sn: The skew-normal and related distributions such as the skew-t and the SUN*. Versioon 2.1.3. Università degli Studi di Padova, Italia. URL: <https://cran.r-project.org/package=sn> (vaadatud 05.05.2026).
- Balanda, K. P. ja H. L. MacGillivray (1988). “Kurtosis: A Critical Review”. *The American Statistician* 42.2, lk. 111–119. URL: <https://doi.org/10.2307/2684482>.
- Eling, M. (2012). “Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?” *Insurance: Mathematics and Economics* 51.2, lk. 239–248. URL: <https://doi.org/10.1016/j.insmatheco.2012.04.001>.
- Hofert, M., I. Kojadinovic, M. Maechler, J. Yan, J. G. Nešlehova ja R. Morger (2026). *Multivariate Dependence with Copulas. Programmable Bibliographies and Citations*. Versioon 1.1.7. URL: <https://cran.r-project.org/web/packages/copula/copula.pdf> (vaadatud 28.04.2026).
- Microsoft (2026). *Microsoft 365 Copilot Chat*. Tehisintellektil põhinev vestlusassistent Microsoft 365 keskkonnas. URL: <https://www.microsoft.com/microsoft-365/copilot> (vaadatud 12.05.2026).
- Overleaf (2026). *Overleaf — Online LaTeX Editor*. Veebipõhine LaTeX-redaktor ja koostöökeskkond. URL: <https://www.overleaf.com> (vaadatud 12.05.2026).
- R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/> (vaadatud 31.04.2026).

- Roon, F. ja P. Karehnke (2017). “A Simple Skewed Distribution with Asset Pricing Applications”. *Review of Finance* 21.6, 2169–2197. URL: <https://doi.org/10.1093/rof/rfw040>.
- Rubio, F. J. (juuni 2013). “Modelling of Kurtosis and Skewness: Bayesian Inference and Distribution Theory”. Doktoritöö. Warwick: University of Warwick.
- Rubio, F. J. ja M. F. J. Steel (2020). “The Family of Two-Piece Distributions”. *Significance* 17.1, lk. 12–13. URL: <https://doi.org/10.1111/j.1740-9713.2020.01352.x>.
- Wallis, K. F. (2014). “The Two-Piece Normal, Binormal, or Double Gaussian Distribution: Its Origin and Rediscoveries”. *Statistical Science* 29.1, lk. 106–112. URL: <http://www.jstor.org/stable/43288461>.
- Winkelbauer, A. (juuli 2014). “Moments and Absolute Moments of the Normal Distribution”. Vienna University of Technology.

Lisa A: Kaheosalise normaaljaotusega andmete genereerimine ja parameetrite hindamine

```
#kaheosalise normaaljaotuse tihedusfunktsioon
dtwopiece_pdf <- function(x, mu, sigma1, sigma2) {
  lamda1 <- 2 * sigma1 / (sigma1 + sigma2)
  lamda2 <- 2 * sigma2 / (sigma1 + sigma2)
  ifelse(x < mu,
    lamda1 * dnorm(x, mean = mu, sd = sigma1),
    lamda2 * dnorm(x, mean = mu, sd = sigma2))
}

library(stats4) #suurima tõepära hinnangu jaoks
library(sn) #Azzalini asümmeetrilise normaaljaotuse pakett

#genereerime mõlemale poole mu-d vastavast pool-normaaljaotusest
rtwopiece <- function(n, mu, sigma1, sigma2) {
  p_vasak <- sigma1 / (sigma1 + sigma2) #vasaku poole osakaal
  u <- runif(n) # genereerime n juhuslikku arvu ühtlasest jaotusest
  #genereerime standardse normaaljaotuse absoluutväärtuse ehk pool-
  #normaaljaotuse ja liidame/lahutame need keskväärtusest
  andmed <- ifelse(u < p_vasak,
    mu - abs(rnorm(n, mean = 0, sd = sigma1)), #vasak pool
    mu + abs(rnorm(n, mean = 0, sd = sigma2))) #parem pool
  return(andmed)
}

set.seed(222) # tulemuste kordamiseks

#genereerime andmed
andmed <- rtwopiece(n = 2000, mu = 0, sigma1 = 0.6, sigma2 = 0.8)

#hindame kaheosalise normaaljaotuse parameetrid
# log-tõepärafunktsioon parameetrite leidmiseks
ll_tpn <- function(mu, sigma1, sigma2) {
  pdf_vals <- dtwopiece_pdf(andmed, mu, sigma1, sigma2)
  -sum(log(pdf_vals))}

```

```

#arvutame väärtused, millest algoritmi alustada
tihedus <- density(andmed)#leiame andmete tiheduse hinnangu
#leiame punkti, kus tihedus on kõige suurem
mood <- tihedus$x[which.max(tihedus$y)]
s1 <- sd(andmed[andmed < mood])
s2 <- sd(andmed[andmed >= mood])
# parameetrite hindamine (mle)
#algväärtus <- list(mu = 0, sigma1 = 0.6, sigma2 = 0.8)
#algväärtus <- list(mu = median(andmed), sigma1 = sd(andmed), sigma2 = sd(
  andmed))
algväärtus <- list(mu = mood, sigma1 = s1, sigma2 = s2)
fit_tpn <- mle(ll_tpn, start = algväärtus,
               method = "L-BFGS-B", #et saaksime lisada alumise piiri ja v
               ältida negatiivseid sigmasid
               lower = c(-Inf, 0, 0))
hinnang_tpn <- coef(fit_tpn)
aic_tpn <- AIC(fit_tpn)
#hindame azzalini asümmeetrilise normaaljaotuse parameetrid, sisseehitatud
  paketi sn abil
fit_sn <- selm(andmed ~ 1, family = "SN")
aic_sn <- AIC(fit_sn)
hinnang_azz <- coef(fit_sn, "DP")
# tavalise normaaljaotuse parameetrid andmete põhjal
mu_n <- mean(andmed)
sd_n <- sd(andmed)
# aic normaaljaotuse jaoks
ll_n <- sum(dnorm(andmed, mean = mu_n, sd = sd_n, log = TRUE))
aic_n <- -2 * ll_n + 2 * 2

```

Lisa B: Logaritmitud kahjuandmetele sobitatud 20 jaotuse AIC sobivusnäitajad

Jaotus	AIC
Kaheosaline normaal-	5 740,38
Normaal-	5 740,44
Asümmeetriline normaal-	5 737,79
Asümmeetriline hüperboolne	5 738,16
Asümmeetriline normaalne Gaussi pöörd-	5 738,12
Asümmeetriline dispersiooni-gamma-	5 738,61
Asümmeetriline üldistatud hüperboolne	5 740,43
Asümmeetriline t-	5 738,06
Asümmeetriline üldistatud hüperboolne	5 740,61
EkspONENT-	8 869,95
Gamma-	6 442,22
Log-normaalne	8 895,12
Logistiline	5 753,92
Weibull	5 923,95
Cauchy	6 264,44
Pareto	8 871,95
Sümmeetriline hüperboolne	5 738,41
Sümmeetriline normaalne Gaussi pöörd-	5 738,38
Sümmeetriline dispersiooni-gamma-	5 738,65
Studenti t-	5 738,12

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Dora Stern,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose “Kaheosalise normaaljaotuse konstrueerimine ja selle sobitamine andmetele”, mille juhendaja on 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'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.

Dora Stern

12.05.2026