



HELSINGIN YLIOPISTO
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATIIKAN JA TILASTOTIETEEN OSASTO

Maisterintutkielma

Keskiylitysfunktion käyttäminen jakauman valinnassa

Tom Huttonin

Tiedekunta/Osasto — Fakultet/Sektion — Faculty		Koulutusohjelma — Utbildningsprogram — Degree programme	
Matemaattis-luonnontieteellinen tiedekunta		Matematiikan aineenopettajan opintosuunta	
Tekijät — Författaren — Authors			
Tom Hutton			
Työn nimi — Arbetets titel — Title			
Keskiylitysfunktion käyttäminen jakauman valinnassa			
Oppiaine — Läroämne — Subject			
Matematiikka			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	
Maisterintutkielma		Maaliskuu 2021	
		Sivumäärä — Sidoantal — Number of pages	
		52 s. + Liitteet 3 s.	
Tiivistelmä — Referat — Abstract			
<p>Joissain käyttökohteissa havaintoaineistoon sovitettavan jakauman hännän paksuuden valinta ei ole täysin suoraviivaista. Jakauman valinta tulee kuitenkin aina perustella jollain tavalla. Kirjallisuudessa alustavan arvion huomiseksi on ehdotettu keskiylitysfunktioiden hyödyntämistä. Keskiylitysfunktiot pohjautuvat vakuutus- ja finanssimatemaattisissa sovelluksissa käytettyihin odotettu tappio-riskimittoihin, joita hyödynnetään sijoitussalkkujen riskillisyyden mittaamisessa. Keskiylitysfunktioiden kuvaajien pohjalta voidaan mallintamiseen valita alustavasti joko kevyt- tai paksuhäntäinen jakauma.</p> <p>Tutkielman tavoitteena on esitellä lukijalle miten keskiylitysfunktioita voidaan käyttää jakauman hännän tyyppin määräämisessä ja mihin matemaattiseen teoriaan keskiylitysfunktioiden käyttö pohjautuu. Tämän lisäksi esitetään keskiylitysfunktioiden ominaisuuksia ja niiden käyttöä tilastollisissa mallinnustilanteissa esimerkkejä hyödyntäen.</p> <p>Tutkielman toisessa luvussa esitetään todennäköisyysteorian perusteisiin kuuluvia määritelmiä ja tutkielmassa käytettäviä merkintätapoja. Kolmannessa luvussa esitetään todistuksista tutkielmassa tarvittavan ääriarvoteorian lauseita ja määritelmiä, joiden päälle keskiylitysfunktioiden käyttämä teoria myöhemmin pohjautuu.</p> <p>Tutkielman neljännessä luvussa keskiylitysfunktiolle esitetään analyyttinen määritelmä ja lasketaan tämän lisäksi joillekin tunnetuille jakaumille keskiylitysfunktion analyyttinen esitysmuoto. Analyytistä määritelmää hyödyntäen todistetaan propositio, joka kertoo millaisia ominaisuuksia keskiylitysfunktiolla on mielivaltaisella epänegatiivisella jakaumalla. Neljännessä luvussa todistetaan myös että jakauman häntäfunktiota voidaan esittää keskiylitysfunktioita käyttämällä ja esitetään kaksi riskienhallinnassa usein käytettyä tapaa havainnollistaa tappioiden vakavuutta: VaR-luku ja odotettu tappio.</p> <p>Tutkielman viidennessä luvussa syvennyttään vakuutus- ja finanssimatemaattisissa sovelluksissa esiintyvien paksuhäntäisten satunnaismuuttujien maailmaan. Luvun lopuksi esitetään tutkielmassa tehtyjen huomioiden sekä kirjallisuuden pohjalta yhteenveto keskiylitysfunktion käyttäytymisen ja jakauman hännän paksuuden välisestä yhteydestä.</p> <p>Tutkielman kuudennessa luvussa esitetään yhteenveto kirjallisuudessa esiintyvistä keskiylitysfunktioiden hyödyntämistavoista. Kirjallisuuskatsauksen muodossa pyritään esittämään lukijalle kattavasti miten keskiylitysfunktion kuvaajia voidaan havaintoaineiston pohjalta piirtää ja mitä piirretyistä kuvaajista voidaan lukea. Luvussa myös sovelletaan tutkielman tuloksia ja kirjallisuuden huomioita kahteen vakuutussovelluksissa kerättyyn aineistoon.</p>			
Avainsanat — Nyckelord — Keywords			
Keskiylitysfunktio, Ääriarvoteoria, Häntäjakauma, Paksuhäntäisyys, Aineistopohjainen päättely			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — Övriga uppgifter — Additional information			

Sisällys

1 Johdanto	1
2 Määritelmiä	3
3 Ääriarvoteoria	5
3.1 Ylitejakaumat	7
3.2 Blokkimaksimimenetelmä	9
3.3 Tason ylittävät huiput -menetelmä	13
4 Keskiylitysfunktiot	16
4.1 Keskiylitysfunktioiden ominaisuuksia	20
4.2 Jakauman määrääminen	23
4.3 Riskimitat	26
5 Paksuhäntäinen keskiylitys	29
5.1 Aliekspontiaaliset jakaumat	30
5.2 Asymptoottinen jakauma	35
5.3 Hännän tyypin määrääminen	39
6 Sovelluksia aineistoihin	41
6.1 Otoskeskiylitysfunktion hyödyntäminen	43
6.2 SOA Group Medical Insurance Large Claims	47
6.3 Norwegian automobile insurance losses	49
7 Lähteet	51
A Liitteet	53
B Tutkielmassa käytetyt R-koodit	55

1 | Johdanto

Havaintoaineistojen pohjalta muodostetaan monenlaisia arvioita aineiston taustalla olevasta satunnaisilmiöstä. Tilastotieteen kursseilla opitaan miten odotusarvoa tai varianssia voidaan arvioida eri tilanteissa, ja hyödyntämällä pelkästään näitä kahta tunnuslukua voidaan muodostaa arvioita aineiston luoneen satunnaismuuttujan jakaumasta. Näitä arvioita hyödyntämällä voidaan tehdä yksinkertaisia arvioita esimerkiksi siitä kuinka suuri nettokertamaksun P tulisi olla tulipalovakuutuksissa, jotta vakuutusyhtiön vakavaraisuus ei vaarantuisi toiminnan aikana.

Esimerkin tilanteessa ongelman muodostaa aineistossa olevat harvinaiset ja erittäin suuret havainnot. Ne ovat selvästi mahdollisia, mutta millä todennäköisyydellä? Tämän todennäköisyyden arvioiminen on hyvä tehdä perusteellisesti, sillä nettokertamaksun P suuruus yksinkertaistetussa tilanteessa voidaan määrittellä ekvivalenssiperiaatteen mukaan muodossa

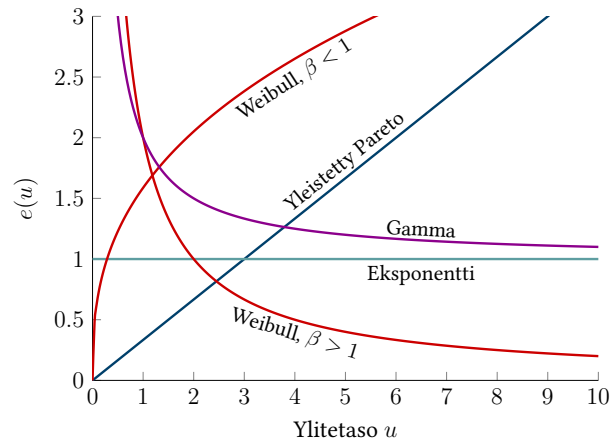
$$P = \mathbb{E}(X) = \mathbb{E}(X \mathbb{1}_{\{X \leq u\}}) + \mathbb{E}(X \mathbb{1}_{\{X > u\}}).$$

Jos tasoa u suurempien havaintojen todennäköisyyttä aliarvioidaan, niin vakuutusyhtiö ei saa kerättyä tarpeeksi vakuutusmaksuja tulevia korvauksia varten. Jos tätä todennäköisyyttä toisaalta yliarvioi, niin tällöin nettokertamaksu P on liian suuri, mikä voi karkoittaa asiakkaita muiden vakuutusyhtiöiden asiakkaita samalla vähentäen yrityksen liikevaihtoa.

Ääriarvoteorian saralla on kehitetty monia eri tapoja arvioida satunnaismuuttujan jakaumaa aineiston pohjalta erittäin suurilla arvoilla. Yksi kirjallisuudessa usein esiintyvä työkalu on *keskiylitysfunktio*, joka määritellään ehdollisen odotusarvon avulla muodossa

$$e(u) = \mathbb{E}(X - u \mid X > u),$$

silloin kun odotusarvo on määritelty. Jokaiseen reaalityyppisellä määriteltyyn integroitavaan todennäköisyysjakaumaan voidaan liittää keskiylitysfunktio, jonka kuvaajia katsomalla voidaan tehdä päätelmiä jakauman tyypistä. Kirjallisuudessa keskiylitysfunktioiden kuvaajia havainnollistetaan usein kuvan 1.1 kaltaisesti.



Kuva 1.1: Yksi tapa havainnollistaa keskiylitysfunktioiden käyttäytymistä eri jakaumien suhteen. Kuvasta voidaan nähdä että havainnollistetut keskiylitysfunktiot voivat olla kasvavia, vakioita tai laskevia. (Hogg & Klugman, 1984; Embrechts ym., 1997)

Tämän tutkielman tavoite on selittää mitä kuvassa 1.1 näkyvä funktioiden käyttäytyminen tarkoittaa ja mihin se matemaattisesti perustuu.

Tutkielman lukijalta odotetaan todennäköisysteorian perusteiden tuntemista, jota tukevia määritelmät esitellään tutkielman toisessa luvussa. Kolmannessa luvussa lukijalle esitetään ääriarvoteorian keskeiset kysymykset ja tulokset sekä esitetään keskiylitysfunktioiden matemaattisen perustan muodostavat lauseet. Näissä kahdessa luvussa ei esitetä todistuksia, vaan niiden tarkoituksena on olla lyhyt yleiskatsaus tässä tutkielmassa tarvittavan ääriarvoteorian maailmaan.

Tutkielman neljännessä luvussa esitetään keskiylitysfunktion määritelmä sekä todistetaan analyttisestä määritelmästä seuraavia ominaisuuksia. Samassa luvussa myös todistetaan, että tiettyjen oletusten ollessa voimassa satunnaismuuttujan jakaumalla ja keskiylitysfunktiolla on yhteys jota voidaan hyödyntää käytettävän jakauman valinnassa. Tutkielman viidennessä luvussa keskitytään paksuhäntäisiin satunnaismuuttujiin ja siihen miten paksuhäntäisyys näkyy keskiylitysfunktion kuvaajassa. Tutkielman kuudennessa luvussa esitetään miten keskiylitysfunktioita voidaan käyttää havaintoaineiston kanssa, mitä kaikkea keskiylitysfunktion kuvaajasta voi päätellä kirjallisuuden mukaan ja sovelletaan kyseisiä kirjallisuuden tuloksia kahteen julkisesti saatavilla olevaan vakuutusyhtiöiden keräämään aineistoon (Dutang & Charpentier, 2020; Reynkens & Verbelen, 2020).

Tutkielman termien suomennoksissa on hyödynnetty lisensointityössä (Leppisaari, 2013) esiintyneitä käännöksiä soveltuvin osin. Englanninkieliset alkuperäistermit on merkitty käännöksen viereen sulkuihin. Tutkielman lähteitä ovat artikkeli (Hall & Wellner, 1981) sekä kirjat (Embrechts ym., 1997; McNeil ym., 2005) ja (Asmussen & Steffensen, 2020).

2 | Määritelmiä

Tässä kappaleessa palautetaan mieleen ja esitellään tutkielmassa tarvittavia määritelmiä ja merkintöjä. Lukijan oletetaan tuntevan todennäköisysteorian perusteet. Määritelmien lähteinä ovat pääasiallisesti (Billingsley, 1995) ja (Grimmett ym., 2001).

Määritelmä 2.1. (Todennäköisyysavaruus) Kolmikkoa $(\Omega, \mathcal{F}, \mathbb{P})$ sanotaan *todennäköisyysavaruudeksi*, jossa

1. Ω on kaikkien mahdollisten tapahtumien avaruus,
2. \mathcal{F} on valittu kokoelma Ω osajoukkoja, jotka muodostavat sigma-algebran ja
3. \mathbb{P} on todennäköisyysmitta, joka antaa jokaiselle joukolle $A \in \mathcal{F}$ arvon väliltä $[0, 1]$.

Määritelmä 2.2. (Satunnaismuuttuja) Kuvausta $X : \Omega \rightarrow \mathbb{R}$ sanotaan *satunnaismuuttujaksi* jos $\{X \in B\} \in \mathcal{F}$ kaikilla $B \in \mathcal{B}$, missä \mathcal{B} on Borelin sigma-algebra reaaliluvuilla.

Määritelmä 2.3. (Indikaattorifunktio) Olkoon $A \subset \Omega$ joukko. *Indikaattorifunktio* $\mathbb{1}_A$ on kuvaus joukolta Ω joukkoon $\{0, 1\}$ siten, että

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$$

kun indikaattori tulkitaan satunnaismuuttujaksi. Satunnaismuuttujien riippuvuus alkeistapahtumasta jätetään usein kirjoittamatta.

Määritelmä 2.4. (Satunnaismuuttujan jakauma) Satunnaismuuttujan X jakauma on määritelty joukon $A \subset \Omega$ suhteen funktiona

$$\mathbb{P}(X \in A).$$

Satunnaismuuttujien X ja Y sanotaan olevan *samoin jakautuneita* jos yhtäsuuruus

$$\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$$

pätee kaikilla $A \subset \Omega$. Satunnaismuuttujien samoin jakautuneisuutta merkitään notaatiolla

$$X \stackrel{d}{=} Y.$$

Määritelmä 2.5. (Triviaali jakauma) Satunnaismuuttujan X jakauman sanotaan olevan *triviaali* jos se saa kaiken todennäköisyysmassansa yhdessä pisteessä $a \in \mathbb{R}$. Triviaalin jakauman kertymäfunktio on muotoa

$$F(x) = \begin{cases} 0, & \text{jos } x < a, \\ 1, & \text{jos } x \geq a. \end{cases}$$

Määritelmä 2.6. (Jakaumasuppeneminen) Jono satunnaismuuttujia X_1, X_2, \dots suppenee jakauman suhteen kohti satunnaismuuttujaa X jos raja-arvo

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

pätee kaikissa kertymäfunktion $F_X(x)$ jatkuvuusasteissa. Jakaumasuppenemistä merkitään

$$X_n \xrightarrow{d} X \quad \text{kun } n \rightarrow \infty.$$

Määritelmä 2.7. (Kertymäfunktio) Funktio $F(x)$ on jonkin satunnaismuuttujan *kertymäfunktio* jos ja vain jos seuraavat neljä ominaisuutta pätevät:

1. $\lim_{x \rightarrow -\infty} F(x) = 0$,
2. $\lim_{x \rightarrow \infty} F(x) = 1$,
3. funktio $F(x)$ on oikealta jatkuva,
4. $F(x)$ on kasvava funktio.

Määritelmä 2.8. (Häntäfunktio) Olkoon $F(x)$ satunnaismuuttujan X kertymäfunktio. Tällöin satunnaismuuttujan X *häntäfunktio* $\bar{F}(x)$ on

$$\bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x).$$

Häntäfunktion ominaisuudet voidaan johdtaa käyttämällä kertymäfunktion ominaisuuksia (Määritelmä 2.7). Kirjallisuudessa satunnaismuuttujan X kuvastaessa elinaikaa häntäfunktiota kutsutaan myös *eloonjäämisfunktiksi* (*Survival function*).

Määritelmä 2.9. (Kantajan päätepiste) Satunnaismuuttujan X *kantajan päätepiste* x_F on piste, jolle pätee

$$x_F = \begin{cases} \inf\{x \mid F(x) = 1\}, & \text{jos } F(x) = 1 \text{ jollain } x \in \mathbb{R}, \\ \infty, & \text{jos } F(x) < 1 \text{ kaikilla } x \in \mathbb{R}. \end{cases}$$

Pisteen x_F voi tulkita olevan viimeinen todennäköisyysmassa omaava piste.

Määritelmä 2.10. (Momentit generoiva funktio) Satunnaismuuttujan X *momentit generoiva funktio* $M_X: \mathbb{R} \rightarrow \mathbb{R}$ on

$$M_X(t) = \mathbb{E}(\exp\{tX\}),$$

joka on äärellinen niissä pisteissä $t \in \mathbb{R}$, joissa odotusarvo $\mathbb{E}(\exp\{tX\})$ on äärellisenä olemassa.

Määritelmä 2.11. (Paksuhäntäinen jakauma) Satunnaismuuttujan X jakauman sanotaan olevan (oikealta) kevythäntäinen, jos

$$M_X(t) < \infty \quad \text{jollain } t > 0.$$

Jos momentit generoiva funktio $M_X(t)$ on ääretön kaikilla $t > 0$, niin tällöin satunnaismuuttuja on (oikealle) paksuhäntäinen.

Määritelmä 2.12. (Hitaasti vaihteleva funktio) Funktiota $L(x): (0, \infty) \rightarrow (0, \infty)$ kutsutaan hitaasti vaihtelevaksi jos kaikilla $\alpha > 0$ pätee

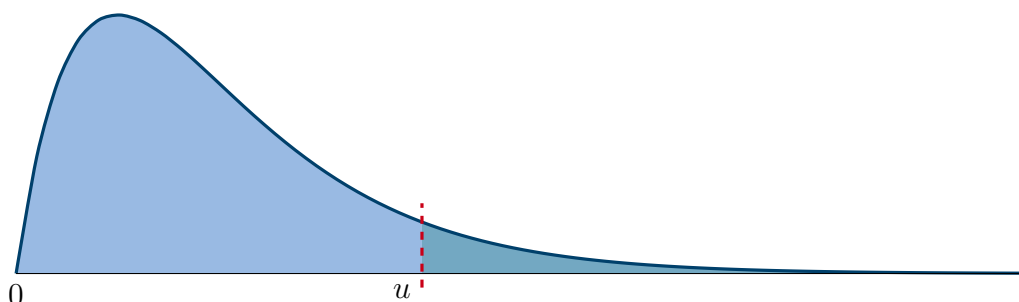
$$\lim_{x \rightarrow \infty} \frac{L(\alpha x)}{L(x)} = 1.$$

Tällöin merkitään $L \in \mathcal{R}_0$.

3 | Ääriarvoteoria

Satunnaismuuttujien muodolle ei voida asettaa yleisiä vaatimuksia, sillä todennäköisyysmassan määrä joukon eri osissa voi vaihdella eri jakaumien välillä. Joidenkin satunnaismuuttujien jakaumat ovat rajoittuneita äärelliselle välille, toiset numeroituvaan määrään pisteitä ja kolmansilla jakauman kantaja on koko reaalilukuakseli. Monien alojen sovelluksissa kiinnostuksen kohteena on useimmiten jakauman oikea häntä. Kun jakauman kantaja on ääretön jompaankumpaan äärettömyyteen, on mahdollista että kyseessä on paksuhäntäinen jakauma (Määritelmä 2.11).

Yksi tapa tarkastella jakauman suurten arvojen käyttäymistä on tutkia jakauman hännän käyttäytymistä tietyn tason $u > 0$ jälkeen. Esimerkiksi vakuutusyhtiöiden näkökulmasta häntätapahtumien arviointia vaaditaan Solvenssi-sääntelyn vuoksi, kun taas rakennustekniikassa padon suunnittelijoita kiinnostaa suurten tulvien aiheuttama vedenpinnan nousu (Embrechts ym., 1997). Tällaisia tapahtumia voidaan tarkastella aluksi kuvaajien kautta.



Kuva 3.1: Gammajakauman tiheysfunktion $f(x)$ kuvaaja ja ylitetaso $u > 0$. Kuvaajassa tummennettu osa kuvastaa tason $u > 0$ ylittävää jakaumaa. Tason $u > 0$ jälkeen jakauman jäljellä olevan todennäköisyysmassan määrä on $\bar{F}(u)$.

Kuvan 3.1 mukaisesti jokainen jakauma on mahdollista jakaa kahteen erilliseen tarkasteltavaan jakaumaan mielivaltaisesta pisteestä $u > 0$ riippuen. Välin $[0, u]$ jakaumaa on havainnollistettu kuvassa 3.1 vaalealla korostuksella ja välin (u, ∞) jakaumaa tummalla korostuksella. Molemmat jakaumat välillä $[0, u]$ ja (u, ∞) ovat hyvin määriteltyjä, jolloin niitä voi käsitellä kuten mitä tahansa muuta jakaumaa, sekä arvioida toisistaan riippumatta.

Tällainen tarkastelu voi olla mielekästä johdannossa esitettyä nettokertamaksua P arvioidessa, jolloin kiinnostuksen kohteena on yhtälö

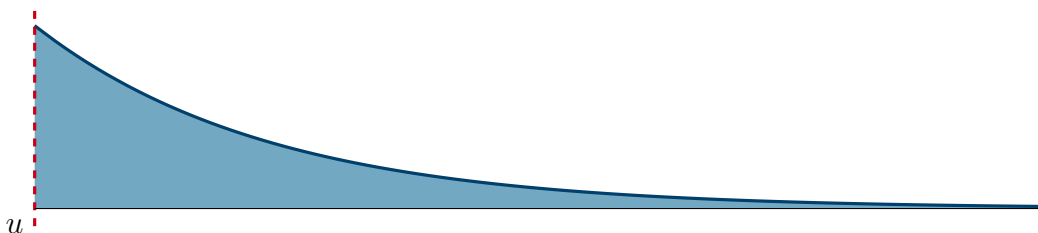
$$P = \mathbb{E}(X) = \mathbb{E}(X \mathbb{1}_{\{0 \leq X \leq u\}}) + \mathbb{E}(X \mathbb{1}_{\{X > u\}}). \quad (3.1)$$

Välin $[0, u]$ jakauma on nyt äärellinen jakauma, jonka odotusarvolle on mahdollista saada

maksimaalinen yläraja-arvio kun $u \geq 1$:

$$\mathbb{E}(X \mathbb{1}_{\{0 \leq X \leq u\}}) = \int_0^u x f(x) dx \leq \max\{x f(x) \mid x \in [0, u]\} \cdot u. \quad (3.2)$$

Yhtälön (3.1) ongelmaksi jääkin väliä (u, ∞) vastaavan odotusarvon suuruuden arviointi. Väliällä $[0, \infty)$ määriteltyjen satunnaismuuttujien tilanteessa kyseisen odotusarvon suuruudelle ei saada yhtälön (3.2) kaltaista välin pituudesta riippuvaa äärellistä yläraja-arviota, sillä välin (u, ∞) pituus on ääretön. Tällöin kyseisen odotusarvon arvioinnissa on hyödynnettävä tietoa häntäfunktion $\bar{F}(u)$ käyttäytymisestä. Häntäfunktion käyttäytymisen havainnollistamisen vuoksi kuvassa 3.2 on piirretty kuvan 3.1 gammajakauman häntä.



Kuva 3.2: Kuvan 3.1 gammajakaumasta saatu ehdollistettu jakauma, kun ehdollistavana tapahtumana on $\{X > u\}$.

Vaikka häntä suppenee jokaisella jakaumalla kohti nollaa, niin pelkästään kuvasta katsoamalla emme saa jakauman hännän vähenemisvauhtia määriteltyä. Kevythäntäisten satunnaismuuttujien häntäfunktion suppenemisnopeudelle on kuitenkin mahdollista löytää eksponentiaalinen yläraja hyödyntämällä momentit generoivan funktion määritelmää 2.10.

Lemma 3.3. *Jos satunnaismuuttujan momentit generoiva funktio on äärellinen jollain $t > 0$, niin tällöin häntäfunktiolla on yläraja*

$$\mathbb{P}(X > x) \leq M_X(t) \exp\{-tx\}.$$

Todistus. Momentit generoivalle funktiolle saadaan indikaattorifunktioita käyttämällä arvioitua alaraja

$$\mathbb{E}(\exp\{tX\}) \geq \mathbb{E}(\exp\{tX\} \mathbb{1}_{\{X > x\}}) \geq \mathbb{E}(\exp\{tx\} \mathbb{1}_{\{X > x\}}). \quad (3.4)$$

Koska $\exp\{tx\}$ on vakio, niin sen voi viedä odotusarvon ulkopuolelle. Tällöin alarajalle saadaan esitys

$$\exp\{tx\} \mathbb{E}(\mathbb{1}_{\{X > x\}}) = \exp\{tx\} \mathbb{P}(X > x),$$

jolloin jakamalla yhtälöketjussa (3.4) positiivinen termi $\exp\{tx\}$ yhtälöiden toiselle puolelle häntäfunktiolle saadaan äärellinen yläraja

$$M_X(t) \exp\{-tx\} \geq \mathbb{P}(X > x).$$

□

Täten jos jakauma on kevythäntäinen, niin kuvan 3.2 häntäfunktion arvioinnissa voidaan käyttää momentit generoivasta funktiosta johdettua eksponentiaalista ylärajaa. Lemma 3.3 ei kuitenkaan anna paksuhäntäisten satunnaismuuttujien häntäfunktioille mitään eksponentiaalista ylärajaa, jolloin häntäfunktion suppenemisnopeudesta ei voida yleisellä tasolla sanoa mitään. Tämä eksponentiaalisen ylärajan puuttuminen häntäfunktiolle on yksi kirjallisuudessa esiintyvä vaihtoehtoinen määritelmä paksuhäntäisille jakaumille.

Ylärajan puuttuminen häntäfunktiolle vaikeuttaa odotusarvon hajotelman (3.1) arviointia, sillä tällöin jälkimmäisen odotusarvon laskemisessa käytetylle integraalille

$$\mathbb{E}(X \mathbb{1}_{\{X > u\}}) = \int_u^\infty x f(x) dx$$

ei saada helposti yläraja-arviota. Tämän vuoksi on pyrittävä muodostamaan erilaisia arviointia häntäfunktion käyttäytymisestä, joita hyödyntämällä voidaan määrätä jakauman tyyppejä ja siitä seuraten myös arvioita yhtälölle (3.1). Häntäfunktion käyttäytymisen tarkka arviointi on kuitenkin vaikeaa pelkästään tiheysfunktioiden kuvaajaa katsomalla, vaan tähän tarkoitukseen tarvitaan parempia työkaluja kuin visuaalinen tarkastelu.

Matematiikan ja tilastotieteen alalla häntäfunktion käyttäytymistä on tarkemmin tarkasteltu äärimmäisten ilmiöiden alalla ja tässä luvussa esitellään pääpiirteittäin tutkielman lukemisessa tarvittavat ääriarvomallinnuksen lauseet todistuksitta. Tämän luvun esitystavassa on mukailtu kirjan (McNeil ym., 2005) lukuja 7.1 ja 7.2.

3.1 | Ylitejakaumat

Kuvissa 3.1 ja 3.2 esiintyneet häntäjakaumat olisi hyvä määritellä myös matemaattisesti, jotta niiden hyödyntäminen olisi perustellumpaa kuin kuvista katsominen. Koska kiinnostuksemme kohdistuu satunnaismuuttujan sellaisiin X arvoihin, joilla pätee $X > u$, niin on luonnollista lähteä tarkastelemaan jakaumaa ehdolla $\{X > u\}$. Tällaisella ehdolla ehdollistettuja jakaumia kutsutaan kirjallisuudessa *ylitejakaumiksi (threshold exceedance)* (Embrechts ym., 1997).

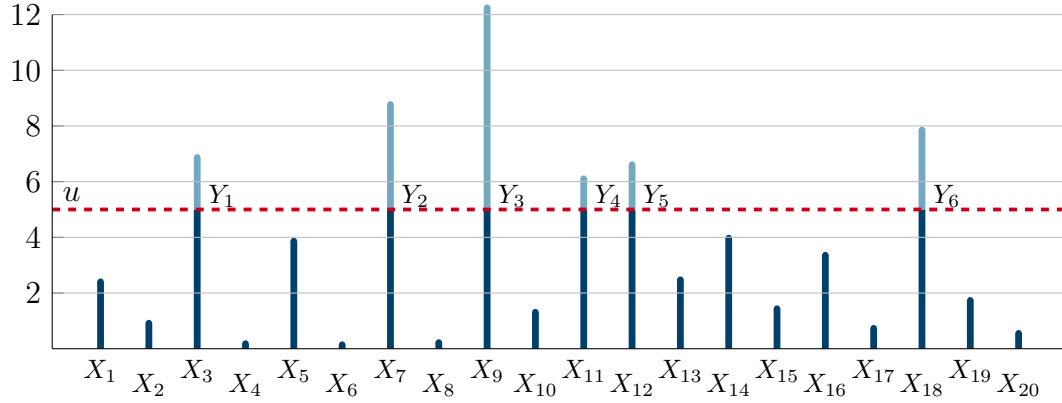
Määritelmä 3.5 (Ylitejakauma). Olkoon $F(x)$ satunnaismuuttujan X kertymäfunktio. Tason u ylittävän jakauman kertymäfunktio $F_u(x)$ on

$$F_u(x) = \frac{F(x+u) - F(u)}{1 - F(u)} = \mathbb{P}(X - u \leq x \mid X > u), \quad \text{kun } x > 0.$$

Helpoin tapa visualisoida ylitejakauma on tarkastella jo edellä esitettyä kuvaa 3.1. Kuvan mukaisesti valitaan taso u , jonka ylittävistä häntäjakaumasta olemme kiinnostuneita. Muodostamalla uusi satunnaismuuttuja Y , jolla on satunnaismuuttujan X ehdollinen jakauma ehdolla $\{X > u\}$, saadaan matemaattisesti määriteltyä tason u ylittävä häntäjakauma. Uutta satunnaismuuttujaa Y voidaan ajatella muodossa

$$Y = X - u, \quad \text{kun } X > u,$$

ja joukkoa $\{X \leq u\}$ vastaavat arvot jätetään huomioimatta. Näin saadaan muodostettua satunnaismuuttuja Y , jonka tulkinta gammajakauman tiheysfunktioita hyödyntämällä on esitetty kuvissa 3.1 ja 3.2. Kuvassa 3.3 on havainnollistettu eksponenttijakaumasta $\text{Exp}(1/5)$ simuloitujen kahdenkymmenen havainnon $\{X_1, \dots, X_{20}\}$ pohjalta lasketut ylitejakauman arvot ylitetasolla $u = 5$.



Kuva 3.3: Eksponenttijakaumasta $\text{Exp}(1/5)$ generoidulle havaintoaineistolle $\{X_1, \dots, X_{20}\}$ ylitetasolla $u = 5$ lasketut ylitejakauman havainnot $\{Y_1, \dots, Y_6\}$.

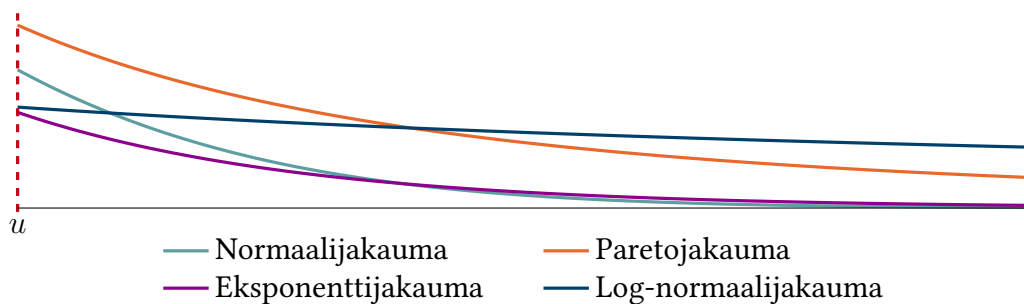
Kuvan 3.3 mukaisesti ylitejakaumaa vastaavia havaintoja $\{Y_1, \dots, Y_6\}$ voidaan käsitellä erillään kaikista havainnoista $\{X_1, \dots, X_{20}\}$, sillä ylitejakauma hyödyntää pelkästään havaintoja $\{Y_1, \dots, Y_6\}$. Täten ehdollistamalla satunnaismuuttuja X ehdolla $\{X > u\}$ saadaan jakauman alkupään vaikutus poistettua tarkastelusta ja jäljelle jää pelkästään jakauman häntä omana positiivisena satunnaismuuttujana Y , jota voidaan käsitellä ja arvioida itsenäisenä satunnaismuuttujana. Erityisesti tälle ehdollistetulle satunnaismuuttujalle voidaan laskea odotusarvo.

Esimerkki 3.6. Tunnetuilla kertymäfunktioilla $F(x)$ ylitejakaumalle on mahdollista saada laskettua yksikäsitteinen esitysmuoto. Kirjallisuudessa klassinen esimerkki tästä on eksponentiaalisesti jakautuneen satunnaismuuttujan ylitejakauma kun jakauman parametri on λ :

$$\begin{aligned}
 F_u(x) &= \frac{F(u+x) - F(u)}{1 - F(u)} \\
 &= \frac{1 - \exp\{-\lambda(u+x)\} - (1 - \exp\{-\lambda u\})}{1 - (1 - \exp\{-\lambda u\})} \\
 &= \frac{-\exp\{-\lambda u\} \exp\{-\lambda x\} + \exp\{-\lambda u\}}{\exp\{-\lambda u\}} \\
 &= 1 - \exp\{-\lambda x\}.
 \end{aligned}$$

Täten eksponenttialajakautuneen satunnaismuuttujan ylitejakauma $F_u(x)$ on myös eksponenttialajakautunut parametrilla λ kaikilla u .

Ylitejakaumia tarkastellessa ollaan kuitenkin koko ajan käsittelemässä satunnaismuuttujien häntiä ylitetasolla $\{u > 0\}$. Tällöin ylitejakaumien arviointia helpottaa jos jakaumien häntien käyttäytymisestä voidaan tehdä yleisiä huomioita. Piirtämällä normaali-, Pareto-, eksponentti-, ja log-normaalijakauman (Määritelmä A.7) hännät samaan kuvaan (Kuva 3.4) on mahdollista huomata esitettyjen häntien olevan laskevia tarpeeksi suurilla u .



Kuva 3.4: Normaali-, eksponentti-, Pareto- ja log-normaali-jakauman hännät pisteestä u eteenpäin.

Voidaanko tästä kuvassa 3.4 näkyvästä häntien laskevuudesta saada pääteltyä jotain yleistä ylitejakaumista ja niiden muodoista? Samaista kysymystä pohdittiin jo 70-luvulla, kunnes ääriarvoteoriaa tutkiessa Pickands (1975) esitti *Pickands–Balkema–de Haan lauseen*. Kyseinen teoreema antaa ylitejakauman kertymäfunktion $F_u(x)$ suppenemiselle kohteen kun u suppenee kohti kantajan päätepistettä x_F . Ennen kyseisen teoreeman esittämistä on kuitenkin tarvetta esitellä kirjallisuudessa olevia ääriarvoteoriaan liittyviä tuloksia ja lisämääritelmiä.

3.2 | Blokkimaksimimenetelmä

Useissa eri käyttökohteissa ääriarvoteoriaa käytetään jakauman maksimin käyttäytymisen arviointiin. Mahdollisia käyttökohteita ovat hydrologiassa sadan tulevan vuoden aikana suurin mahdollinen meren vedenpinnan nousu, vakuutuslalla kymmenen vuoden aikana suurin mahdollinen korvaus tai geologiassa maanjäristyksen kovin mahdollinen voimakkuus 50 vuoden aikana. (Asmussen & Steffensen, 2020) Näissä sovelluskohteissa tärkeintä on saada muodostettua jonkinlainen arvio maksimijakaumalle $M_n = \max\{X_1, X_2, \dots, X_n\}$, missä havaintojen X_1, X_2, \dots, X_n oletetaan olevan riippumattomia ja samoin jakautuneita. Havaintojen X_1, \dots, X_n riippumattomuusoletusta käyttämällä maksimin kertymäfunktio voidaan esittää muodossa

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \stackrel{\text{iid}}{=} \mathbb{P}(X_1 \leq x) \cdots \mathbb{P}(X_n \leq x) = F^n(x). \quad (3.7)$$

Tämänlainen n havainnon maksimin tarkastelu muodostaa klassisen ääriarvoteorian perustan. Kirjallisuudessa tätä lähestymistapaa nimitetään *blokkimaksimimenetelmäksi* (*Block maxima method*) ja maksimia M_n *blokkimaksimiksi* (*Block maxima*). Blokkimaksimimenetelmässä satunnaismuuttujan X realisaatioista otetaan n ensimmäistä havaintoa X_1, \dots, X_n , joiden sanotaan muodostavan oman blokkinsa. Blokin voi tulkita sisältävän esimerkiksi yhden kuukauden, yhden vuoden tai sadan vuoden aikana realisoituneita havaintoja satunnaismuuttujasta X . Tarkastelemalla miten blokkia vastaava blokkimaksimi M_n käyttäytyy kun n kasvaa saadaan tietoa blokkimaksimin asymptoottisesta käyttäytymisestä. (McNeil ym., 2005)

Maksimin kertymäfunktion M_n käyttäminen arvioinnissa voi kuitenkin aiheuttaa ongelmia otoskoon kasvaessa tarpeeksi suureksi, kuten on havainnollistettu esimerkissä 3.8.

Esimerkki 3.8. Olkoon satunnaismuuttuja X tasajakautunut välillä $[0, 1]$. Nyt jokaisella $x \in [0, 1)$ pätee $F^n(x) \rightarrow 0$ kun $n \rightarrow \infty$, kun taas valinnalla $x = 1$ suppenemiselle pätee $F^n(1) \rightarrow 1$. Merkitään kertymäfunktion $F^n(x)$ suppenemisen kohteena olevaa kertymäfunktioita $G(x)$, jolloin $F^n(x) \rightarrow G(x)$ kun $n \rightarrow \infty$. Kertymäfunktio $G(x)$ on täten muotoa

$$G(x) = \begin{cases} 0, & x \in [0, 1), \\ 1, & x = 1, \end{cases}$$

jonka kantaja on alkuperäisen välin $[0, 1]$ sijaan pistejoukko $\{0, 1\}$.

Esimerkin 3.8 kaltaisesta pistekeskittyneestä jakaumasta ei ole paljon hyötyä sovelluskohteissa. Tämän ongelman välttämiseksi kirjallisuudessa tarkastellaan maksimin M_n affiinia muutosta

$$M_n^* = \frac{M_n - b_n}{a_n}, \quad (3.9)$$

missä $a_n > 0$ ja $b_n \in \mathbb{R}$ ovat sopivia normittavia lukujonoja (McNeil ym., 2005). Muunnoksella saatua termiä M_n^* kutsutaan normitetuksi maksimiksi. Normitetun maksimin käyttäytymistä tarkastelemalla on mahdollista saada muodostettua tarkempi arvio maksimijakaumasta, sillä affini muunnos tuo paremmin esille maksimin asymptoottisen käyttäytymisen (Asmussen & Steffensen, 2020). Sopivien normittavien lukujonojen selvittämisen matemaattista taustaa on esitetty kirjallisuudessa (Embrechts ym., 1997). Eri jakaumille tarvitaan eri jonot, eikä kaikille jakaumille F voida välttämättä löytää sopivia jonoja lainkaan.

Mistä tiedämme, että tällaisen affiinin muutoksen tarkastelu ei tuota arvioinnissa ongelmia? Yhtälöä (3.9) käyttämällä normitettu maksimi saadaan esitettyä muodossa

$$M_n^* = a_n^{-1}M_n - b_n/a_n = \alpha_n M_n - \beta_n, \quad (3.10)$$

missä $\alpha_n = a_n^{-1} > 0$ ja $\beta_n = b_n/a_n \in \mathbb{R}$. Skaalaamalla maksimia tällä tavoin pyritään välttämään esimerkin 3.8 kaltainen tilanne, missä rajalla $n \rightarrow \infty$ maksimilla M_n olisi triviaali jakauma. Tästä esitysmuodosta on selkeämpää nähdä, että normitetun ja normaalin maksimin välinen yhteys riippuu pelkästään vakioiden α_n ja β_n arvoista. Kun kyseisten vakioiden arvot on selvitetty, niin tällöin määritelmän A.1 mukaan normitettu maksimi ja normaali maksimi ovat *samantyyppisiä jakaumia* (Embrechts ym., 1997), jolloin voimme merkitä

$$M_n^* \stackrel{d}{=} \alpha_n M_n - \beta_n.$$

Koska jakaumat M_n ja M_n^* ovat vain skaalattuja ja siirrettyjä versioita toisistaan, niin tällöin maksimin M_n jakauman arviointiin riittää pelkästään normitetun maksimin M_n^* käyttäytymisen tarkastelu. Koska normitetulla maksimilla saamme asymptoottisen käyttäytymisen paremmin mallinnettua, niin tällöin myös maksimijakauman asymptoottinen käyttäytyminen saadaan tarkemmin arvioitua aineiston pohjalta.

Normitetun maksimin arvojen laskeminen on myös suoraviivaista yhtälöä (3.7) hyödyntämällä. Kun $n \geq 2$ on kiinnitetty, niin tällöin normitetun maksimin häntäfunktion arvo pisteessä x on

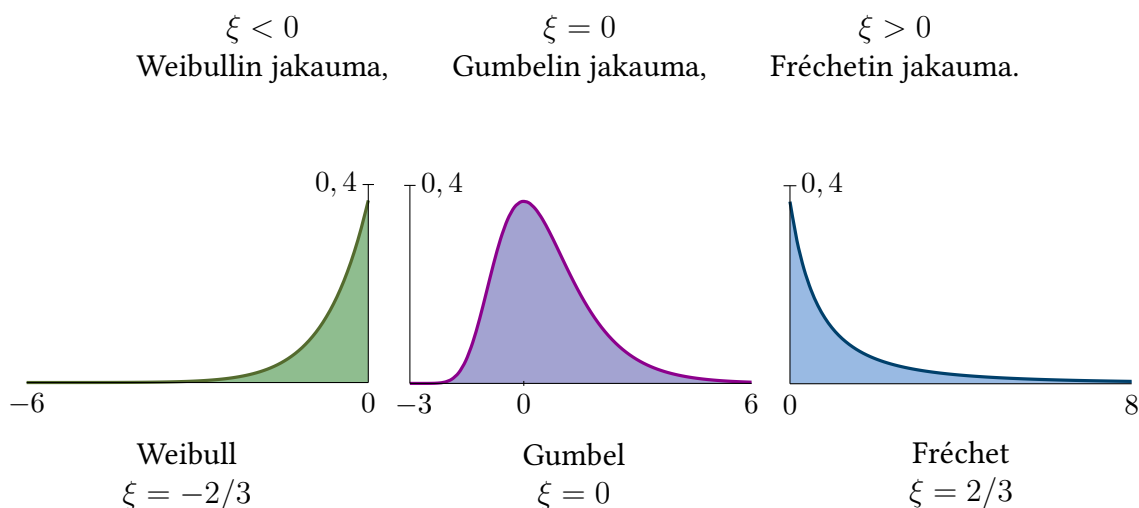
$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \mathbb{P}(M_n \leq a_n x + b_n) = F^n(a_n x + b_n).$$

Täten normitetun maksimin kertymäfunktion arvojen laskemiseksi kiinnitetyllä $n \geq 2$ tarvitaan pelkästään satunnaismuuttujien X_1, \dots, X_n kertymäfunktiot sekä vakioiden a_n ja b_n arvot. (Embrechts ym., 1997)

Määritelmä 3.11. (Yleistetty ääriarvojakauma) Yleistetyn ääriarvojakauman $H_\xi(x)$ (*Generalized extreme value distribution*) kertymäfunktio on

$$H_\xi(x) = \begin{cases} \exp\{-(1 + \xi x)^{-1/\xi}\}, & \xi \neq 0, \\ \exp\{-e^{-x}\}, & \xi = 0, \end{cases}$$

missä $1 + \xi x > 0$. Parametrin ξ arvo määrää kertymäfunktion $H_\xi(x)$ tyyppin yhteen seuraavasta kolmesta tyyppistä:



Kuva 3.5: Määritelmän 3.11 mukaisen Weibullin, Gumbelin ja Fréchetin ääriarvojakaumien tiheysfunktiot origon ympäristössä.

Parametrin ξ arvon määräämät jakaumatyyppit kuvaavat kuvan 3.5 mukaisesti kolmea erilaista todennäköisyysjakaumaa. Gumbelin jakauma on määritelty kaikilla reaaliluvuilla, Fréchetin jakauma on määritelty pelkästään positiivisilla ja Weibullin jakauma pelkästään negatiivisilla reaaliluvuilla. Weibullin ääriarvojakaumasta puhuttaessa tarkoitetaan oikeastaan negatiivista Weibullin jakaumaa, jonka kantaja on positiivisten reaalilukujen sijaan negatiiviset reaaliluvut. Ääriarvojakaumista puhuttaessa Weibullin jakauma on vakiintunut termi jakaumatyypille kun $\xi < 0$. (Asmussen & Steffensen, 2020)

On myös hyvä huomata, että Gumbelin, Fréchetin ja Weibullin jakaumaperheiden unioni ei sisällä kaikkia mahdollisia jakaumia. Esimerkiksi Poissonin sekä geometrinen jakauma eivät kuulu ääriarvojakaumien perheeseen (Asmussen & Steffensen, 2020). Kirjallisuudessa on kuitenkin osoitettu, että oleellisesti kaikki vakuutusmatemaattisissa sovelluskohteissa esiintyvät jatkuvat jakaumat kuuluvat johonkin edellä mainituista ääriarvojakaumaperheistä (Hogg & Klugman, 1984; McNeil ym., 2005). Täten Gumbelin, Fréchetin ja Weibullin ääriarvojakaumien käyttö on perusteltua mallinnettaessa riippumattomien havaintojen pohjalta jakaumien suuria arvoja (Embrechts ym., 1997). Tarkempaa tietoa esiteltujen ääriarvojakaumien ominaisuuksista löytyy kirjallisuudesta (Embrechts ym., 1997; McNeil ym., 2005).

Käyttökohteita erikseen silmälläpitäen olisi hyödyllistä tietää miten normitettu maksimi käyttäytyy suurilla n . Tämän kysymyksen vastausta, *Fisher-Tippnett-Gnedenkon lausetta*, kutsutaan kirjallisuudessa ääriarvoteorian ensimmäiseksi lauseeksi.

Lause 3.12. (*Fisher–Tippett–Gnedenkon lause*) Jos on olemassa lukujonot $a_n > 0$ ja $b_n \in \mathbb{R}$ joilla suppeneminen

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq x \right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x) \quad (3.13)$$

kohti epätriviaalin jakauman kertymäfunktioita $H(x)$ pätee, niin tällöin kertymäfunktio $H(x)$ kuuluu johonkin määritelmän 3.11 mukaisista ääriarvojakaumaperheistä.

Todistus. Todistus löytyy kirjasta (Asmussen & Steffensen, 2020, luku IX:4). \square

Täten jos normitetun maksimin kertymäfunktio suppenee kohti jotakin kertymäfunktioita $H(x)$, niin tällöin kertymäfunktion $H(x)$ on oltava joko Fréchetin, Gumbelin tai Weibullin jakauman tyyppinen. Vakioiden a_n ja b_n arvojen lopullisesta valinnasta riippuen on mahdollista saada rajankäynnillä $n \rightarrow \infty$ erilainen ääriarvojakauman tyyppiin määräävä parametrin ξ arvo kuin muilla normittavilla vakioilla.

Vaikka kertymäfunktio $H(x)$ saa rajalla $n \rightarrow \infty$ vakioiden a_n ja b_n valinnoista riippuen hieman eri arvon, niin yhtälön (3.13) mukaisen suppenemisen pätiessä *tyypittäisen suppenemisen lause* (Lause A.2) mukaan suppenemisen kohteena olevat kertymäfunktiot ovat samaa tyyppiä. (Embrechts ym., 1997) Normittavat vakiot on aina mahdollista valita siten, että rajalla $n \rightarrow \infty$ kertymäfunktio $H(x)$ on samalainen kuin määritelmässä 3.11 oleva kertymäfunktio $H_\xi(x)$ (McNeil ym., 2005).

Tästä tiedosta emme saa suoraan lisätietoa yksittäisen kertymäfunktion $F(x)$ muodosta. Lauseen 3.12 antama suppenemistulos ei ole täysin hyödytön, sillä kertymäfunktiot $F(x)$ ja $H(x)$ eivät ole täysin irrallisia toisistaan. Seuraava määritelmä antaakin meille yhteyden näiden kertymäfunktioiden välille.

Määritelmä 3.14. (Jakauman vaikutuspiiri maksimin suhteen) Kun yhtälössä (3.13) oleva suppeneminen pätee, niin tällöin jakaumasuppenemisen perusteella pätee

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} H(x),$$

missä $H(x)$ on yksi määritelmässä 3.11 esiintyvistä ääriarvojakaumista. Tällöin kertymäfunktion $F(x)$ sanotaan kuuluvan *jakauman $H(x)$ maksimin vaikutuspiiriin* (*Maximal domain of attraction, MDA*), jota merkitään $F \in MDA(H)$.

Siitä, että yhtälö (3.13) pätee saadaan myös tietää, että tarpeeksi suurilla n kertymäfunktio $F^n(x)$ alkaa muistuttamaan yhtä määritelmän 3.11 mukaisia ääriarvojakaumia. Täten äärimmäisiä arvoja mallinnettaessa onkin perusteltua käyttää Weibullin, Gumbelin tai Fréchetin tyyppisiä jakaumia. Tarkempaa tarkastelua jakauman vaikutuspiiristä maksimin suhteen sekä ohjeita miten valita oikea ääriarvotyyppi löytyy kirjallisuudesta (Asmussen & Steffensen, 2020, luku IX:3).

Vaikka blokkimaksimimenetelmä on pitkälle jalostettu tapa mallintaa satunnaismuuttujien äärimmäisiä arvoja, niin sillä on kuitenkin omat heikkoutensa. Suurin heikkous liittyy havaintoaineiston käyttöön; blokkimaksimi käyttää mallinnuksessa pelkästään blokin suurinta arvoa ja jättää muut mahdolliset äärimmäiset arvot huomiotta (McNeil ym., 2005). Maksimia pienemmät äärimmäiset arvot kuitenkin kertovat hyödyllistä tietoa jakauman häntäkäyttäytymisestä, sillä niiden avulla voidaan arvioida jakauman hännän paksuutta.

3.3 | Tason ylittävät huiput -menetelmä

Kirjallisuudessa blokkimaksimimenetelmän lisäksi äärimmäisiä arvoja mallinnetaan nykyaikaisemmalla *tason ylittävät huiput -menetelmällä* (*Peaks-over-Threshold, POT*). Tason ylittävät huiput -menetelmää käyttäessä valitaan taso $u > 0$ jonka ylittävästä jakaumasta ollaan kiinnostuneita. Kaikista kyseisen tason ylittäneistä havainnoista X_i ($i \in \{1, \dots, n\}$) vähennetään taso u , jolloin alkuperäisestä havainnosta jää jäljelle arvo $X_i - u > 0$. Tarkastelun kohteena on täten samanlainen tilanne kuin määritelmän 3.5 mukaisessa ylitejakaumassa.

Tason ylittävät huiput -menetelmä täten säilyttää tiedon kaikista tason u ylittävistä tapahtumista, ja siitä kuinka paljon taso u ylitettiin. Äärimmäisten ilmiöiden harvinaisuuden vuoksi havaintoja häntätapahtumista on vähän, ja tason ylittävät huiput -menetelmän avulla kaikkia äärimmäisiä havaintoja hyödynnetään mallinnuksessa. Tällaista lähestymistapaa pidetään kirjallisuudessa hyvänä valintana ääriarvoteoriaa soveltaessa, sillä tällöin saadaan käytettyä mahdollisimman tehokkaasti koko havaintoaineisto X_1, \dots, X_n . (McNeil ym., 2005)

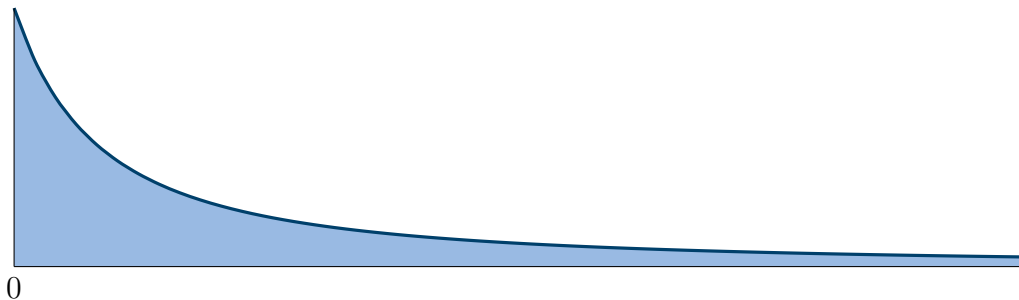
Määritelmä 3.15. (Yleistetty Pareto-jakauma) Yleistetyn Pareto -jakauman (*Generalized Pareto Distribution, GPD*) kertymäfunktio on muotoa

$$G_{\xi, \beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\{-x/\beta\}, & \xi = 0, \end{cases}$$

missä $\beta > 0$ ja $x \geq 0$ kun $\xi \geq 0$ ja $0 \leq x \leq -\beta/\xi$ kun $\xi < 0$. Parametrit ξ ja β ovat jakauman muoto- ja asteikkoparametri. Parametri ξ määrää hännän paksuuden, jonka vuoksi jakauman momenttien olemassaolo vaatii tarpeeksi pienen arvon kyseiselle parametrille. Odotusarvo ja varianssi ovat

$$\mathbb{E}(X) = \frac{\beta}{1 - \xi} \quad (\text{kun } \xi < 1) \quad \text{ja} \quad \text{Var}(X) = \frac{\beta^2}{(1 - \xi)^2(1 - 2\xi)} \quad (\text{kun } \xi < 1/2).$$

Yleistetyn Pareto-jakauman hyödyllisyys ylitejakaumien arvioinnissa on mahdollista nähdä kuvia 3.4 ja 3.6 vertailemalla. Jokainen kuvassa 3.4 näkyvä häntä on laskeva, jonka perusteella voisi päätellä kuvitteellisen yleisen häntäjakauman olevan laskeva suurilla u . Laskevuuteen pohjautuen yleistetty Pareto-jakauma on hyvä ehdokas yleiseksi häntäjakaumaksi, sillä sen tiheysfunktio on aidosti laskeva. Etuna on myös jakauman määrittely kahta eri parametria (ξ ja β) käyttäen, joka mahdollistaa jakauman tarkemman hienosäädön verrattuna yksiparametriisiin jakaumiin.



Kuva 3.6: Yleistetyn pareto-jakauman tiheysfunktio (parametrein $\xi = 2$ ja $\beta = 1$)

Valitsemalla parametrit ξ ja β sopivasti on mahdollista saada esitettyä muiden tunnettujen jakaumien kertymäfunktioita yleistetyn Pareto-jakauman kertymäfunktion avulla (McNeil ym., 2005). Valinnoilla $\xi = 0$ ja $\beta = 1$ kertymäfunktioiksi saadaan

$$G_{\xi,\beta}(x) = G_{0,1}(x) = 1 - \exp\{-x\},$$

joka on eksponenttijakauman kertymäfunktio parametrilla 1. Yleistetyn Pareto-jakauman määrittelyjoukkoa $[0, \infty)$ muokkaamalla on mahdollista esittää normaali Pareto-jakauma yleistettynä Pareto-jakaumana. Muuntamalla argumentti x muotoon $\frac{x-\mu}{\beta}$, missä parametria μ sanotaan sijaintiparametriksi, saadaan uudeksi määrittelyjoukoksi $x \in [\mu, \infty)$ ja kertymäfunktioiksi

$$G_{\xi,\beta,\mu}(x) = 1 - \left(1 + \frac{\xi(x-\mu)}{\beta}\right)^{-1/\xi}. \quad (3.16)$$

Valitsemalla tässä esitysmuodossa sijaintiparametriksi $\mu = \beta/\xi$ saamme kertymäfunktioiksi

$$G_{\xi,\beta,\mu}(x) = 1 - \left(1 + \frac{\xi(x - \beta/\xi)}{\beta}\right)^{-1/\xi} = 1 - \left(\frac{\xi/\beta}{x}\right)^{1/\xi},$$

joka on Pareto-jakauman kertymäfunktio parametrein $\alpha = 1/\xi$ ja $k = \xi/\beta$ (Määritelmä A.4). Yhtälössä (3.16) käytetty kolmiparametrinen (ξ, β, μ) esitysmuoto on vaihtoehtoinen määritelmä yleistetyille Pareto-jakaumalle. Tässä työssä kuitenkin keskitymme tilanteeseen $\mu = 0$.

Lisämääritelmien esittelyn jälkeen voimme vihdoin palata takaisin luvun alussa esitettyyn kysymykseen: onko olemassa jotain kuvassa 3.4 esiintyvää häntäjakaumien yleistä käyttäytymistä takaavaa tulosta? Seuraava *Picklands-Balkema-De Haanin lause* antaa meille käyttökelpoisen tuloksen ylitejakaumien tyyppin määrittelemisessä.

Lause 3.17. (*Pickands–Balkema–de Haanin lause*) Olkoon x_F jakauman F kantajan päätepiste, $G_{\xi,\beta(u)}(x)$ yleistetyn Pareto-jakauman kertymäfunktio parametrein ξ ja $\beta(u)$ ja $F_u(x)$ jakauman F ylitejakauma tasolla u . On mahdollista löytää positiivisarvoinen funktio $\beta(u)$ siten, että suppeneminen

$$\lim_{u \rightarrow x_F} \sup_{x \in [0, x_F - u)} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0,$$

pätee jos ja vain jos $F \in MDA(H_\xi)$, $\xi \in \mathbb{R}$.

Todistus. Lauseen todistus on esitetty artikkelissa (Pickands, 1975). □

Lausetta 3.17 kutsutaan kirjallisuudessa ääriarvoteorian toiseksi lauseeksi. Nimitys pohjautuu siihen, että blokkimaksimimenetelmän sijasta lause pohjautuu tason ylittävät huiput -menetelmään. Käytännössä Pickands–Balkema–de Haanin lauseen mukaan tason u kasvaessa tarpeeksi suureksi ylitejakauman $F_u(x)$ ja yleistetyn Pareto-jakauman $G_{\xi,\beta(u)}(x)$ kertymäfunktioit ovat tietyssä mielessä lähellä toisiaan. Tällöin ylitejakauman arvioinnissa voidaan käyttää arviota

$$F_u(x) \approx G_{\xi,\beta(u)}(x) \quad (3.18)$$

tarpeeksi suurilla u . Yllä olevaa approksimaatiota hyödyntämällä ylitejakauman arviointi yksinkertaistuu yleistetyn Pareto-jakauman parametrien arvioinniksi. (Embrechts ym., 1997)

Kertymäfunktion $G_{\xi, \beta(u)}(x)$ parametrien ξ ja $\beta(u)$ estimaattoreiden $\hat{\xi}$ ja $\hat{\beta}(u)$ arvioiminen tehdään sovellustilanteissa kerätyn aineiston pohjalta. On hyvä huomata, että $\beta(u)$ on muuttujan u funktio, eikä kiinnitetty vakio kuten ξ . Käytännössä lausetta 3.17 sovellettaessa ainestopohjaisessa mallinnuksessa onkin tarvetta käyttää tarpeeksi suurta tasoa u , jotta arviota (3.18) voisi perustellusti käyttää. Kirjallisuudessa (Embrechts ym., 1997, luku 6.5.1) ja (McNeil ym., 2005, luku 7.2.2) käydään läpi miten ylitejakauman parametreja voidaan arvioida eri menetelmiä hyödyntämällä.

Ääriarvoteoriasta on mahdollista lukea kattavempi tarkastelu lähteistä (Embrechts ym., 1997), (McNeil ym., 2005, luku 7) ja (Asmussen & Steffensen, 2020, luku IX).

4 | Keskiylitysfunktiot

Sovelluskohteissa kiinnostuksen kohteena on usein arvioitavien tapahtumien odotusarvo. Käyttöyhteydestä riippuen odotusarvo voi tarjota hyödyllistä tietoa esimerkiksi odotettavissa olevasta liikennemäärästä, viikon sademäärästä tai tulevista tappioista sijoitustoiminnassa. Normaalia odotusarvoa laskettaessa laskeminen tapahtuu satunnaismuuttujan koko määrittelyjoukon yli, mutta tämä ei enää käy jos kiinnostuksen kohteena ovat pelkästään satunnaismuuttujan häntätapahtumat.

Yksi tapa rajoittaa satunnaismuuttujan häntään on tarkastella tason ylittävät huiput - menetelmän mukaisesti tason u ylittäviä tapahtumia $\{X > u\}$. Edellisessä luvussa näytettiin, että kyseisellä ehdollistuksella saadut kertymäfunktiot $F_u(x)$ määrittelevät hyvin satunnaismuuttujien jakaumia. Koska ylitejakaumat ovat pohjimmiltaan jakaumia, niin niille on mahdollista laskea odotusarvo ehdollistamalla alkuperäinen satunnaismuuttuja X tapahtumalla $\{X > u\}$.

Tason u ei myöskään tarvitse olla koko ajan sama kiinteä luku. Esimerkiksi kaupungin hulevesijärjestelmää suunnitellessa voi olla mielekästä ottaa huomioon vielä satamattoman sademäärän muutokset kun vettä on satanut 10 mm tai 60 mm. Ehdollistettu odotusarvo oletettavasti saa kummankin sademäärän kohdalla eli arvon, joka kertoo kuinka paljon lisää vettä on taivaalta tulossa kun vettä on satanut jo tietyn verran. Hulevesijärjestelmiä suunnitellessa kyseisiä lukuja käyttämällä saadaan järjestelmä mitoitettua paremmin myös poikkeuksellisen suuria sademääriä varten.

Koska taso u voi saada monia eri arvoja, niin on mielekästä tulkita se funktion argumentiksi. Tällaisella lähestymistavalla päädytään *keskiylitysfunktioiden* määritelmään:

Määritelmä 4.1 (Keskiylitysfunktio). Olkoon $\bar{F}(u)$ aidosti positiivisen satunnaismuuttujan X jakauman häntäfunktio siten, että $\bar{F}(0) = 1$. Tällöin keskiylitysfunktio (*mean excess function*) on

$$e(u) = \mathbb{E}(X - u \mid X > u), \quad \text{kun } u \geq 0,$$

ja $e(u) = 0$ kun $\bar{F}(u) = 0$.

Lemma 4.2. *Keskiylitysfunktion analyyttinen muoto on*

$$e(u) = \frac{1}{\bar{F}(u)} \int_u^\infty \bar{F}(x) \, dx.$$

Todistus. Olkoon X aidosti positiivinen satunnaismuuttuja. Tällöin keskiylitysfunktion määritelmää käyttämällä saadaan

$$\begin{aligned}\mathbb{E}(X - u \mid X > u) &= \frac{\mathbb{E}((X - u)\mathbb{1}_{\{X > u\}})}{\mathbb{P}(X > u)} \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty (x - u)f(x) \, dx.\end{aligned}$$

Koska

$$-f(x) = \frac{d}{dx}(1 - F(x)) = \frac{d}{dx}\bar{F}(x),$$

niin edellisen yhtälön integraali voidaan osittaisintegroinnilla esittää muodossa

$$\begin{aligned}\int_u^\infty (x - u)f(x) \, dx &= \int_u^\infty (u - x)(-f(x)) \, dx \\ &= \lim_{t \rightarrow \infty} \int_u^t (u - x)\bar{F}(x) \, dx + \int_u^\infty \bar{F}(x) \, dx \\ &= \lim_{t \rightarrow \infty} (u\bar{F}(t) - t\bar{F}(t)) + \int_u^\infty \bar{F}(x) \, dx \\ &= -\lim_{t \rightarrow \infty} t\bar{F}(t) + \int_u^\infty \bar{F}(x) \, dx.\end{aligned}\tag{4.3}$$

Odotusarvon määritelmästä osittaisintegroinnilla saadaan

$$\begin{aligned}\mathbb{E}(X) &= -\int_0^\infty x(-f(x)) \, dx \\ &= -\lim_{t \rightarrow \infty} \int_0^t x\bar{F}(x) \, dx + \int_0^\infty \bar{F}(x) \, dx \\ &= -\lim_{t \rightarrow \infty} t\bar{F}(t) + \int_0^\infty \bar{F}(x) \, dx.\end{aligned}\tag{4.4}$$

Koska epänegatiivisen satunnaismuuttujan odotusarvo saadaan tunnetusti yhtälöstä

$$\mathbb{E}(X) = \int_0^\infty \bar{F}(x) \, dx,$$

niin yhtälöstä (4.4) seuraa yhtäsuuruus

$$-\lim_{t \rightarrow \infty} t\bar{F}(t) = 0,$$

joka sijoittamalla yhtälöön (4.3) saadaan yhtäsuuruus

$$\int_u^\infty (x - u)f(x) \, dx = \int_u^\infty \bar{F}(x) \, dx,$$

todistaen lemmän väitteen. □

Keskiylitysfunktiolle voidaan käyttää eri käyttöyhteyksissä eri nimityksiä. Luotettavuusteoriassa keskiylitysfunktio esiintyy samalla nimellä kuin tässä työssä. Henkivakuutusmatematiikan yhteydessä funktiota $e(u)$ kutsutaan *jäljellä olevan elinajan odotusarvoksi* (*mean residual life*). (Hall & Wellner, 1981)

Esimerkki 4.5. Keskiylitysfunktiolle on tunnettujen kertymäfunktioiden tapauksessa mahdollista laskea analyyttinen esitysmuoto. Olkoon satunnaismuuttuja X eksponenttijakautunut parametrilla λ . Tällöin keskiylitysfunktio $e(u)$ on muotoa:

$$\begin{aligned}
 e(u) &= \mathbb{E}(X - u \mid X > u) \\
 &= \frac{1}{\bar{F}(u)} \int_u^\infty \bar{F}(x) \, dx \\
 &= \frac{1}{\exp\{-\lambda u\}} \int_u^\infty \exp\{-\lambda x\} \, dx \\
 &= -\frac{1}{\lambda \exp\{-\lambda u\}} \int_u^\infty -\lambda \exp\{-\lambda x\} \, dx \\
 &= \frac{1}{\lambda \exp\{-\lambda u\}} \exp\{-\lambda u\} \\
 &= \frac{1}{\lambda},
 \end{aligned}$$

joka tukee edellisen luvun esimerkissä 3.6 laskettua tulosta eksponenttijakauman ylitejakaumasta. Koska ylitejakauma on eksponenttijakautunut parametrilla λ , niin ylitejakauman odotusarvo on eksponenttijakauman mukainen $1/\lambda$.

Esimerkki 4.6. Yksi kirjallisuudessa usein käytetty keskiylitysfunktio on yleistetyn Pareto-jakauman keskiylitysfunktio. Olkoon satunnaismuuttujalla X yleistetty Pareto-jakauma parametrein $\xi \in (0, 1)$ ja $\beta > 0$. Tällöin keskiylitysfunktio on muotoa

$$\begin{aligned}
 e(u) &= \frac{1}{\left(1 + \frac{\xi u}{\beta}\right)^{-1/\xi}} \int_u^\infty \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} \, dx \\
 &= \frac{1}{\left(1 + \frac{\xi u}{\beta}\right)^{-1/\xi}} \frac{1}{\frac{\xi}{\beta}} \int_u^\infty \frac{\xi}{\beta} \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} \, dx \\
 &= \frac{1}{\left(1 + \frac{\xi u}{\beta}\right)^{-1/\xi}} \frac{\left(1 + \frac{\xi u}{\beta}\right)^{-1/\xi+1}}{-\frac{\xi}{\beta} \left(-\frac{1}{\xi} + 1\right)} \\
 &= \frac{\frac{1}{\beta}(\beta + \xi u)}{\frac{1}{\beta}(1 - \xi)} \\
 &= \frac{\beta + \xi u}{1 - \xi},
 \end{aligned}$$

kun $u \geq 0$. Vastaavalla tavalla tilanteelle $\xi < 0$ saadaan sama esitysmuoto keskiylitysfunktiolle määrittelyjoukolla $0 \leq u \leq -\beta/\xi$ (Embrechts ym., 1997).

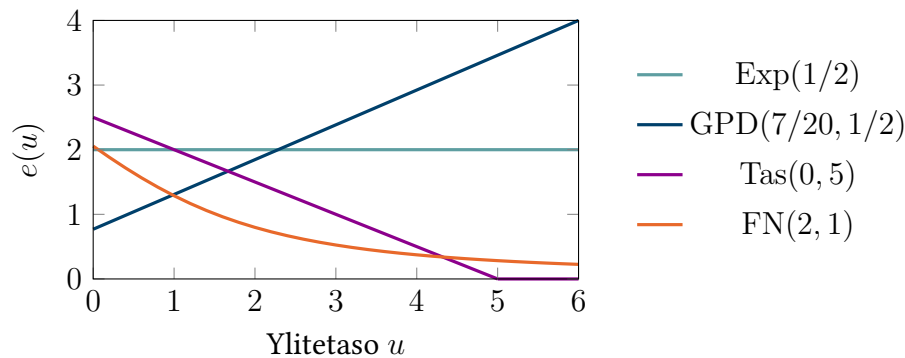
Esimerkki 4.7. Olkoon satunnaismuuttuja X tasajakautunut välillä $[0, a]$, missä $a \in (0, \infty)$. Tällöin välillä $[0, a]$ keskiylitysfunktio on muotoa

$$\begin{aligned} e(u) &= \frac{1}{1 - u/a} \int_u^a 1 - \frac{x}{a} dx \\ &= \frac{1}{(a - u)/a} \frac{(a - u)^2}{2a} \\ &= \frac{a - u}{2}, \end{aligned}$$

ja muulloin $e(u) = 0$.

Vaikka edellä esitetyissä esimerkeissä on onnistuttu laskemaan keskiylitysfunktiolle analyyttinen esitysmuoto, niin kaikilla jakaumilla se ei ole mahdollista, sillä lemmassa 4.2 olevaa integraalia ei välttämättä voida laskea suljetussa muodossa. Tarvittaessa keskiylitysfunktiolle voidaan tietokoneohjelmilla laskea numeerinen likiarvo käyttäen tunnettua häntäfunktiota $\bar{F}(x)$. Kuvassa 4.1 on esimerkin vuoksi piirretty numeerisesti laskettu (Koodi B.1) taitetulle normaalijakaumalle $\text{FN}(2, 1)$ (Määritelmä A.5) keskiylitysfunktio.

Analyttisesti tai numeerisesti ratkaistuja keskiylitysfunktioita voidaan havainnollistaa piirtämällä niiden kuvaajat. Tällöin visuaalisten havaintojen perusteella on helpompi vertailla keskiylitysfunktioiden käyttäytymistä.



Kuva 4.1: Esimerkeissä 4.5–4.7 laskettujen sekä numeerisesti ratkaistu taitetun normaalijakauman $\text{FN}(2, 1)$ keskiylitysfunktioiden kuvaajat. Jakaumien parametrit on valittu keskiylitysfunktion käyttäytymisen havainnollistamiseksi.

Kuvassa 4.1 on piirretty esimerkeissä 4.5–4.7 laskettujen keskiylitysfunktioiden kuvaajat sekä numeerisesti laskettu taitetun normaalijakauman keskiylitysfunktion kuvaaja. Näistä kuvaajista voidaan havaita, että niin tasa- kuin taitetun normaalijakauman tilanteessa keskiylitysfunktio on laskeva, kun taas eksponenttijakauman tapauksessa kuvaaja on vakio ja yleistetyyn Pareto-jakauman tapauksessa kuvaaja on kasvava. Näistä jakaumista tasa-, taitettu normaali- ja eksponenttijakauma ovat kevythäntäisiä, kun taas yleistetty Pareto-jakauma on paksuhäntäinen.

4.1 | Keskiylitysfunktioiden ominaisuuksia

Häntäfunktioiden käyttäytymiselle tunnetaan yleisiä tuloksia vaikka konkreettista suljetua muotoa ei tunnetaisi. Häntäfunktiota käyttämällä voidaan laskea satunnaismuuttujan odotusarvo ja sitä käyttämällä voidaan laskea jakauman hännän suppenemisnopeudelle arvioita. Voidaanko näitä tunnettuja ominaisuuksia käyttämällä saada johdettua funktion $e(u)$ käyttäytymiselle yleisiä tuloksia, jotka eivät riipu taustalla vaikuttavasta aidosti positiivisesta satunnaismuuttujasta X ?

Keskiylitysfunktioita käsittelevässä artikkelissa (Hall & Wellner, 1981) on esitetty hyödyllisiä tuloksia kun satunnaismuuttujan X kertymäfunktion $F(x)$ oletetaan olevan oikealta jatkuva ja odotusarvon $\mathbb{E}(X) = \mu$ oletetaan olevan äärellinen. Seuraava propositio 4.8 ja sen todistus on esitetty alkuperäistä artikkelia mukailleen.

Propositio 4.8. (Keskiylitysfunktion ominaisuudet) Oletetaan, että aidosti positiivisen satunnaismuuttujan X kertymäfunktio on oikealta jatkuva, piste x_F on jakauman F kantajan päätepiste ja $\mathbb{E}(X) = \mu < \infty$. Tällöin keskiylitysfunktiolla on seuraavat ominaisuudet:

- i. $e(u)$ on epänegatiivinen ja oikealta jatkuva.
- ii. $e(0) = \mu > 0$.
- iii. Funktio $v(u) = e(u) + u$ on kasvava.
- iv. Keskiylitysfunktiolla on vasemmat raja-arvot välillä $(0, \infty)$ ja epäjatkuvuuskohdissa se hyppää ylöspäin.
- v. Jos $x_F < \infty$, niin
 - (a) $\lim_{x \rightarrow u^-} e(x) > 0$ kaikilla $u \in (0, x_F)$,
 - (b) $\lim_{x \rightarrow x_F^-} e(x) = 0$,
 - (c) $e(u)$ on jatkuva pisteessä x_F .
- vi. $\bar{F}(u) = \frac{e(0)}{e(u)} \exp\left\{-\int_0^u \frac{1}{e(x)} dx\right\}$.
- vii. $\int_0^u \frac{1}{e(x)} dx \rightarrow \infty$ kun $u \rightarrow x_F$.

Todistus. Näytetään proposition 4.8 jokainen väite todeksi väite kerrallaan. Luettavuuden parantamiseksi todistuksessa toispuolisista raja-arvoista käytetään lyhennysmerkintöjä

$$\lim_{x \rightarrow t^+} f(x) = f(t+) \quad \text{ja} \quad \lim_{x \rightarrow t^-} f(x) = f(t-). \quad (4.9)$$

- i. Lemman 4.2 antamasta esitysmuodosta

$$e(u) = \frac{1}{\bar{F}(u)} \int_u^\infty \bar{F}(x) dx$$

on mahdollista nähdä, että $1/\bar{F}(u)$ ja $\int_u^\infty \bar{F}(x) dx$ ovat aina epänegatiivisia, josta seuraa myös $e(u)$ epänegatiivisuus. Koska integraalifunktio $i(u) = \int_u^\infty \bar{F}(x) dx$ on jatkuva, erityisesti oikealta jatkuva ja termi $1/\bar{F}(u)$ on oikealta jatkuva, niin tällöin myös $e(u)$ on oikealta jatkuva.

- ii. Asettamalla $u = 0$ lemmän 4.2 antamaan esitysmuotoon saadaan

$$e(0) = \frac{1}{\bar{F}(0)} \int_0^\infty \bar{F}(x) dx = \int_0^\infty \bar{F}(x) dx = \mathbb{E}(X) = \mu.$$

Odotusarvon μ positiivisuus seuraa oletetusta satunnaismuuttujan X aidosti positiivisesta jakaumasta.

- iii. Erotusta $v(u+t) - v(u)$ tarkastelemalla on mahdollista löytää epänegatiivinen alaraja funktiolle $v(u)$ kaikilla $t > 0$. Häntäfunktion vähenevyyden perusteella funktion $v(u)$ määritelmästä saadaan alaraja

$$\begin{aligned}
 v(u+t) - v(u) &= \frac{1}{\bar{F}(u+t)} \int_{u+t}^{\infty} \bar{F}(x) \, dx + u + t - \left(\frac{1}{\bar{F}(u)} \int_u^{\infty} \bar{F}(x) \, dx + u \right) \\
 &= \frac{1}{\bar{F}(u+t)} \int_{u+t}^{\infty} \bar{F}(x) \, dx - \frac{1}{\bar{F}(u)} \int_u^{\infty} \bar{F}(x) \, dx + t \\
 &\geq \frac{1}{\bar{F}(u)} \left(\int_{u+t}^{\infty} \bar{F}(x) \, dx - \int_u^{\infty} \bar{F}(x) \, dx \right) + t \\
 &= -\frac{1}{\bar{F}(u)} \int_u^{u+t} \bar{F}(x) \, dx + t. \tag{4.10}
 \end{aligned}$$

Yhtäsuuruutta (4.10) hyödyntämällä on mahdollista saada erotuksen alarajaksi 0. Kirjoittamalla tämä väite epäyhtälömuodossa päädyimme tarkastelemaan epäyhtälöä

$$t\bar{F}(u) \geq \int_u^{u+t} \bar{F}(x) \, dx.$$

Määrätyn integraalin perusominaisuuksiin vetoamalla epäyhtälön yhtäsuuruus pätee pelkästään silloin kun $\bar{F}(x) = \bar{F}(u)$ välillä $[u, u+t]$. Muulloin häntäfunktion vähenevyydestä johtuen epäyhtälön vasen puoli on suurempi kuin yhtälön oikea puoli. Täten erotuksen $v(u+t) - v(u)$ alaraja on 0, josta seuraa funktion $v(u)$ kasvavuus.

- iv. Koska häntäfunktiolla on vasemmat raja-arvot välillä $(0, \infty)$, niin täten termillä $1/\bar{F}(u)$ ja integraalifunktiolla $i(u) = \int_u^{\infty} \bar{F}(x) \, dx$ on vasemmat raja-arvot välillä $(0, \infty)$. Täten funktiolla $e(u)$ on vasemmat raja-arvot kaikkialla.

Olkoon piste t funktion $e(u)$ epäjatkuvuuskohta. Erotus $e(t+) - e(t-)$ voidaan kirjoittaa muodossa

$$e(t+) - e(t-) = e(t+) + t - e(t-) - t = v(t+) - v(t-) \geq 0,$$

missä viimeisen termin positiivisuus seuraa kohdassa iii. näytetystä funktion $v(u)$ kasvavuudesta.

- v. a) Kun $u \in (0, x_F)$, niin tällöin

$$e(u) = \frac{1}{\bar{F}(u)} \int_u^{\infty} \bar{F}(x) \, dx = \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x) \, dx > 0,$$

jolloin vasemmat raja-arvot $e(u-)$ ovat positiivisia kaikilla u .

- b) Keskiylitysfunktion määritelmästä saamme $e(x_F) = 0$. Koska funktio $v(u)$ on kasvava, niin epäyhtälö $v(u) \leq v(x_F)$ pätee kaikilla $u \leq x_F$. Tästä epäyhtälöstä saamme keskiylitysfunktiolle ylärajan

$$e(u) \leq x_F - u.$$

Kun $u \rightarrow x_F-$, niin keskiylitysfunktion arvon ylärajaksi pisteessä x_F tulee 0. Koska $e(u)$ on myös epänegatiivinen proposition ensimmäisen kohdan mukaan, niin tällöin on oltava $e(x_F-) = 0$.

- c) Koska funktio $e(u)$ on aina oikealta jatkuva, niin tällöin jatkuvuuden näyttämiseksi riittää varmistaa funktion toispuoleisten raja-arvojen yhtäsuuruus pisteessä x_F . Koska $e(x_F) = 0$ määritelmän mukaan, niin täten $e(x_F+) = 0$. Kohdan b) mukaan $e(x_F-) = 0$, joten

$$e(x_F-) = 0 = e(x_F+),$$

josta saamme pääteltyä funktion $e(u)$ jatkuvuuden pisteessä x_F .

- vi. Koska odotusarvon μ oletettiin olevan äärellinen, niin tällöin epänegatiivisen satunnaismuuttujan odotusarvon kaavan avulla voidaan näyttää, että epäyhtälö

$$\mathbb{E}(X) = \int_0^\infty \bar{F}(t) dt \geq \int_x^\infty \bar{F}(t) dt \quad (4.11)$$

pätee kaikille $x \in [0, \infty)$. Koska yhtälön (4.11) viimeinen integraali on äärellinen, niin täten häntäfunktion arvo pisteessä x voidaan ilmaista muodossa

$$\bar{F}(x) = -\frac{d}{dx} \int_x^\infty \bar{F}(t) dt.$$

Edellistä häntäfunktion $\bar{F}(x)$ esitysmuotoa hyödyntämällä integraali $\int_0^u \frac{1}{e(x)} dx$ voidaan ilmaista muodossa

$$\begin{aligned} \int_0^u \frac{1}{e(x)} dx &= \int_0^u \frac{\bar{F}(x)}{\int_x^\infty \bar{F}(t) dt} dx \\ &= -\int_0^u \frac{\frac{d}{dx} \int_x^\infty \bar{F}(t) dt}{\int_x^\infty \bar{F}(t) dt} dx \\ &= -\int_0^u d \log \left(\int_x^\infty \bar{F}(t) dt \right) \\ &= -\int_0^u d \log(\bar{F}(x)e(x)) \\ &= -[\log(\bar{F}(u)e(u)) - \log(\bar{F}(0)e(0))] \\ &= -\log \left(\bar{F}(u) \frac{e(u)}{e(0)} \right), \end{aligned}$$

josta seuraa yhtäsuuruus

$$\int_0^u \frac{1}{e(x)} dx = -\log \left(\bar{F}(u) \frac{e(u)}{e(0)} \right) \quad (4.12)$$

mistä $\bar{F}(u)$ voidaan ratkaista ottamalla molemmanpuoleiset eksponentit, jolloin saadaan esitys

$$\bar{F}(u) = \frac{e(0)}{e(u)} \exp \left\{ -\int_0^u \frac{1}{e(x)} dx \right\}. \quad (4.13)$$

vii. Käyttämällä yhtälöä (4.12) integraalille $\int_0^u \frac{1}{e(x)} dx$ saadaan esitysmuoto

$$\int_0^u \frac{1}{e(x)} dx = -\log\left(\frac{\bar{F}(u) \bar{F}(u)^{-1} \int_u^\infty \bar{F}(t) dt}{e(0)}\right) = -\log\left(\frac{\int_u^\infty \bar{F}(t) dt}{e(0)}\right),$$

missä logaritmin argumentissa $\int_u^\infty \bar{F}(t) dt \rightarrow 0$ kun $u \rightarrow x_F$. Tästä seuraa suppeneminen $\int_0^u \frac{1}{e(x)} dx \rightarrow \infty$ kun $u \rightarrow x_F$.

□

Täten olettamalla aidosti positiivisen satunnaismuuttujan kertymäfunktion olevan oikealta jatkuva sekä odotusarvon olevan äärellinen saadaan johdettua keskiylitysfunktiolle jo verrattaen yleisiä tuloksia. Johdetut tulokset eivät riipu tarkemmin taustalla vaikuttavasta jakaumasta, vaan pätevät kaikille oletukset täyttävälle jakaumille.

Keskiylitysfunktion määritelmästä saimme pääteltyä että $e(u)$ saa epänegatiivisia arvoja kaikilla $u \geq 0$. Erityistapauksena saatiin myös $e(0) = \mu$, joka sitoo keskiylitysfunktion arvoja ja jakauman odotusarvoja lähemmäs toisiaan. Nämä kaksi esimerkkiä myöskin havainnollistavat ylitejakauman odotusarvon käyttäytymistä muuttujan u saadessa eri arvoja.

Ongelmia ei myöskään ole luvassa jakauman kantajan päätepisteen x_F ympäristössä. Jos kyseinen piste on äärellinen, niin tällöin proposition 4.8 viidennen kohdan mukaan keskiylitysfunktio on kyseisessä pisteessä jatkuva. Täten esimerkiksi välin $[0, 5]$ tasajakauman keskiylitysfunktiolle pätee $e(u) = 0$ kun $u \geq 5$ kuvan 4.1 mukaisesti.

Yhteenvedona proposition 4.8 mukaan satunnaismuuttujan täyttäessä proposition oletukset keskiylitysfunktion käyttäytymisessä ei ole luvassa mitään yllättävää. Tämän säännönmukaisen käyttäytymisen vuoksi keskiylitysfunktion voi tulkita säännönmukaisesti havainnollistavan eri jakaumien ylitejakauman käyttäytymistä. Tällöin varsinkin proposition kuudennen kohdan antaman häntäfunktion esitysmuodon (4.13) perusteella kuvien 1.1 ja 4.1 kaltaisista kuvaajista voidaan tehdä päätelmiä keskiylitysfunktioiden taustalla vaikuttavista jakaumista.

4.2 | Jakauman määrittäminen

Edellisessä luvussa esitetyn proposition 4.8 kuudennen kohdan mukaan häntäfunktio $\bar{F}(u)$ voidaan esittää keskiylitysfunktiosta riippuvassa muodossa. Mitä ominaisuuksia funktiolta $e(u): [0, \infty) \rightarrow [0, \infty)$ pitää kuitenkin olettaa, jotta yhtälö (4.13) määräisi aidosti positiivisen satunnaismuuttujan häntäfunktion? Seuraavassa *Esityslauseessa* esitetään vähimmäisolelutukset keskiylitysfunktiolle, jolla saatu esitysmuoto toteuttaa häntäfunktiolta vaaditut ominaisuudet. Lause 4.14 sekä sen todistus on esitetty alunperin artikkelissa (Hall & Wellner, 1981), jota mukaillen lause todistetaan tässä tutkielmassa lyhennysmerkintöjä (4.9) hyödyntäen.

Lause 4.14. (Esityslause) Oletetaan, että funktiolla $e(u) : [0, \infty) \rightarrow [0, \infty)$ on seuraavat ominaisuudet:

1. $e(u)$ on oikealta jatkuva ja $e(0) > 0$,
2. $v(u) = e(u) + u$ on ei-laskeva,
3. jos $e(u-) = 0$ jollain $u = x_0$, niin tällöin $e(u) = 0$ kun $u \in [x_0, \infty)$,
4. jos $e(u-) > 0$ kaikilla u , niin tällöin $\int_0^\infty \frac{1}{e(u)} du = \infty$.

Olkoon $x_T = \inf\{x : e(x-) = 0\} \leq \infty$. Määritellään funktio

$$\bar{F}(x) = \begin{cases} \frac{e(0)}{e(x)} \exp\left\{-\int_0^x \frac{1}{e(t)} dt\right\}, & 0 \leq x < x_T, \\ 0, & x \geq x_T, \end{cases} \quad (4.15)$$

kuten proposition 4.8 kuudennessa kohdassa.

Tällöin $F(x) = 1 - \bar{F}(x)$ on kertymäfunktio välillä $[0, \infty)$, jolle pätee $F(0) = 0$ ja $x_F = x_T$. Kertymäfunktion $F(x)$ määräämällä jakaumalla on tällöin myös äärellinen odotusarvo $\mu_F = e(0) < \infty$ sekä muotoa $e_F(u) = e(u)$ oleva keskiylitysfunktio.

Todistus. Jotta kertymäfunktio $F(x)$ olisi hyvin määritelty aidosti positiivisen jakauman kertymäfunktio, niin tällöin yhtälössä $F(x) = 1 - \bar{F}(x)$ olevan häntäfunktion tulee toteuttaa seuraavat ehdot:

1. $\bar{F}(x)$ on epänegatiivinen, oikealta jatkuva ja $\bar{F}(0) = 1$,
2. $\bar{F}(x)$ on laskeva,
3. $\bar{F}(x) > 0$ kun $x < x_T$. Jos $x_T = \infty$ niin $\bar{F}(x) \rightarrow 0$ kun $x \rightarrow \infty$,
4. $\int_0^\infty \bar{F}(u) du = \mu_F < \infty$.

Kohtien 1–3 vaatimuksilla varmistetaan että annettu funktio (4.15) toteuttaa häntäfunktiolta vaaditut ominaisuudet. Neljäs vaatimus varmistaa sen, että keskiylitysfunktio on äärellisenä määritelty pisteessä 0. Näyttämällä nämä neljä kohtaa saadaan lauseen väite todistettua.

1. Proposition 4.8 ensimmäisen kohdan mukaan keskiylitysfunktio $e(u)$ on epänegatiivinen ja oikealta jatkuva. Koska häntäfunktio on yhtälön (4.15) mukaan näiden muunnos, niin tällöin myös häntäfunktio on oikealta jatkuva ja epänegatiivinen.

Asettamalla $x = 0$ häntäfunktion yhtälöön (4.15) saadaan

$$\bar{F}(0) = \frac{e(0)}{e(0)} \exp\left\{-\int_0^0 \frac{1}{e(t)} dt\right\} = 1.$$

2. Häntäfunktion laskevuuden osoittamiseksi tarkastellaan osamäärää $\bar{F}(x+t)/\bar{F}(x)$ kun $0 < t \leq t+x < x_T$:

$$\frac{\bar{F}(x+t)}{\bar{F}(x)} = \frac{e(x)}{e(x+t)} \exp\left\{-\int_x^{x+t} \frac{1}{e(u)} du\right\}. \quad (4.16)$$

Jotta häntäfunktio olisi laskeva, niin yllä olevan osamäärän tulisi olla pienempi kuin yksi. Integraalille $\int_x^{x+t} \frac{1}{e(u)} du$ on mahdollista saada alaraja

$$\begin{aligned}\int_x^{x+t} \frac{1}{e(u)} du &= \int_x^{x+t} \frac{1}{v(u) - u} du \\ &\geq \int_x^{x+t} \frac{1}{v(x+t) - u} du \\ &\geq \log\left(\frac{e(x+t) + t}{e(x+t)}\right).\end{aligned}$$

Logaritmin argumenttia arvioimalla saadaan integraalille pienempi alaraja

$$\log\left(\frac{e(x+t) + t}{e(x+t)}\right) = \log\left(\frac{v(x+t) - x}{e(x+t)}\right) \geq \log\left(\frac{v(x) - x}{e(x+t)}\right) = \log\left(\frac{e(x)}{e(x+t)}\right),$$

joka sijoittamalla yhtälöön (4.16) antaa häntäfunktioiden osamäärälle ylärajan

$$\frac{\bar{F}(x+t)}{\bar{F}(x)} \leq \frac{e(x)}{e(x+t)} \exp\left\{-\log\left(\frac{e(x)}{e(x+t)}\right)\right\} = \frac{e(x)}{e(x+t)} \frac{e(x+t)}{e(x)} = 1.$$

Saadusta ylärajasta voidaan nähdä $\bar{F}(x+t) \leq \bar{F}(x)$, joka osoittaa häntäfunktioiden laskevuuden.

3. Välillä $0 < x < x_T < \infty$ keskiylitysfunktion $e(u)$ arvoille on olemassa alaraja $\varepsilon > 0$, jolle pätee $\inf\{e(t) : 0 < t < x\} \geq \varepsilon$. Täten keskiylitysfunktio (4.15) on positiivinen välillä $(0, x)$, jonka vuoksi osamäärä $e(0)/e(t)$ on positiivinen kaikilla $t \in (0, x)$. Häntäfunktiossa $\bar{F}(x)$ olevalle integraalille pätee täten $\int_0^x \frac{1}{e(u)} du < \infty$, joten häntäfunktio on positiivinen välillä $(0, x)$.

Häntäfunktion $\bar{F}(x)$ rajakäyttäytyminen $x \rightarrow x_T$ tilanteessa $x_T = \infty$ voidaan näyttää käyttämällä integraalille $\int_x^{x+t} \frac{1}{e(u)} du$ edellisessä kohdassa saatua alarajaa $\log\left(\frac{e(x+t)+t}{e(x+t)}\right)$ valinnalla $x = 0$, jolloin pätee

$$\int_0^t \frac{1}{e(u)} du \geq \log\left(\frac{e(t) + t}{e(t)}\right).$$

Käyttämällä kyseistä arviota häntäfunktiolle saadaan yläraja

$$\bar{F}(t) \leq \frac{e(0)}{e(t)} \exp\left\{-\log\left(\frac{e(t) + t}{e(t)}\right)\right\} = \frac{e(0)}{e(t)} \frac{e(t)}{e(t) + t} = \frac{e(0)}{v(t)} \rightarrow 0,$$

kun $t \rightarrow \infty$. Täten häntäfunktiolle pätee $\bar{F}(x) \rightarrow 0$ kun $x \rightarrow \infty$.

4. Odotusarvon μ laskemisessa voidaan käyttää muotoa $\mathbb{E}(X) = \int_0^\infty \bar{F}(u) du$, sillä oletuksen mukaan satunnaismuuttuja X on aidosti positiivinen. Integraalin $\int \bar{F}(u) du$ ratkaisu saadaan derivoimalla yhtäsuuruus (4.12) muuttujan u suhteen, jolloin saamme yhtäsuuruuden

$$\frac{1}{e(u)} = -\frac{d}{du} \log(\bar{F}(u)e(u)) = -\frac{\frac{d}{du} \bar{F}(u)e(u)}{\bar{F}(u)e(u)},$$

josta häntäfunktiolle saadaan ratkaistua esitysmuoto

$$\bar{F}(u) = -\frac{d}{du} (\bar{F}(u)e(u)).$$

Sijoittamalla saatu esitysmuoto odotusarvon kaavaan saadaan

$$\begin{aligned}\mathbb{E}(X) &= \int_0^\infty \bar{F}(u) \, du = - \int_0^{x_T} \frac{d}{du} (\bar{F}(u)e(u)) \, du \\ &= - \int_0^{x_T} \bar{F}(u)e(u) \, du = - (\bar{F}(x_T)e(x_T) - \bar{F}(0)e(0)) = e(0) = \mu < \infty.\end{aligned}$$

Koska yhtälön (4.15) mukainen häntäfunktio $\bar{F}(x)$ toteuttaa ehdot 1–4, niin täten se myös määrää jakauman positiivisilla reaalityyppisillä. \square

Täten keskiylitysfunktiota käyttämällä on mahdollista saada esitettyä jakauman häntäfunktio $\bar{F}(x)$. Koska häntäfunktio määrää yksikäsitteisesti satunnaismuuttujan jakauman, niin täten keskiylitysfunktioiden avulla on mahdollista määrittellä häntäfunktion kautta satunnaismuuttujan jakauma. Myöhemmin tätä yhteyttä tullaan käyttämään aineistopohjaisen päättelyn tukena.

Käytännön ongelmaksi muodostuu kuitenkin keskiylitysfunktion $e(x)$ suljetun esitysmuodon saaminen. Aineistopohjaisissa ongelmissa ei välttämättä ole mahdollista esittää keskiylitysfunktiota analyttisen häntäfunktion esitysmuodon puuttuessa, vaan on turvaututtava aineistopohjaisiin arvioihin keskiylitysfunktioista. Luvussa 6 esiteltävän empiirisen keskiylitysfunktion (Määritelmä 6.1) avulla on mahdollista saada laskettua aineistopohjainen keskiylitysfunktio.

On myös huomautettava, että todistuksessa oletettiin satunnaismuuttujan odotusarvon μ_F olevan äärellinen. Häntäfunktion määritelmä itsessään ei kuitenkaan vaadi odotusarvon äärellisyyttä, sillä häntäfunktiolta vaaditaan yleisessä tilanteessa pelkästään ehtojen 1.–3. toteutuminen. Tämä rajoittava oletus johtuu häntäfunktion esitysmuodossa (4.15) esiintyvistä luvusta $e(0)$, joka proposition 4.8 toisen kohdan perusteella on yhtä kuin odotusarvo μ_F . Jos odotusarvon sallittaisiin olevan ääretön, niin tällöin häntäfunktio olisi ääretön kaikilla x .

4.3 | Riskimitat

Rahoitukseen ja vakuutukseen liittyvissä sovelluskohteissa yksi kiinnostuksen kohde on usein häntätapahtumien suuruusluokka jonkin ennalta määrätyn aikavälin Δ aikana. Vahinkovakuutuksessa voidaan olla kiinnostuneita miljoonan euron rajapyykin ylittävien tappioiden kokonaistappiomäärästä kymmenen vuoden aikana ja sijoitustoiminnan matemaattikassa osakesalkun maksimitappiosta yksittäisen viikon aikana. Yksi tapa havainnollistaa suuria tappiota aikavälillä Δ olisi pyrkiä laskemaan niiden vakavuutta kuvastavia tunnuslukuja. Esimerkkejä tunnusluvuista ovat odotusarvo $\mathbb{E}(X)$ ja varianssi $\text{Var}(X)$, jotka kertovat jotain satunnaismuuttujan X jakaumasta. Näiden lisäksi keskiylitysfunktiot ovat yksi tapa havainnollistaa häntätapahtumien suuruusluokkaa.

Keskiylitysfunktioiden lisäksi kirjallisuudessa on käytössä muita työkaluja miten tällaisia satunnaismuuttujan X häntän aiheuttamia riskejä mallinnetaan. Tällaisia työkaluja kutsutaan *riskimitoiksi*. Riskimittojen idea on kuvastaa yrityksen kohtaamaa riskiä yksittäisen tunnusluvun kautta, jota voidaan käyttää myös yrityksen sisäisessä päätöksenteossa (Asmussen & Steffensen, 2020). Yleisessä käytössä on kaksi eri riskimittoa: *VaR-luku* ja *odotettu tappio* (McNeil ym., 2005; Asmussen & Steffensen, 2020). Kumpikin näistä antaa käyttäjälleen lisätietoa häntätapahtuman vakavuudesta yksittäisen tunnusluvun kautta, joita hyö-

dyntämällä saadaan muodostettua tarkempi kuva mahdollisten tappioiden suuruusluokasta ja täten myös yrityksen kyvystä selvitä kyseisistä tappioista. Usein kyseisistä tunnusluvuista ovat kiinnostuneita myös finanssialaa valvovat viranomaiset, jotka haluavat tunnuslukujen pysyvän säätelyn määrämien rajojen sisäpuolella (Asmussen & Steffensen, 2020).

Määritelmä 4.17. (VaR-luku) Olkoon X tappioita kuvaava satunnaismuuttuja. VaR-luku (*Value-at-Risk*) luottamustasolla $\alpha \in (0, 1)$ kuvaa tappion vähimmäissuuruutta ylemmässä α -kvantiilissa, joka saadaan laskettua kaavalla

$$VaR_\alpha(X) = \inf\{t \in \mathbb{R} \mid \mathbb{P}(X \leq t) \geq \alpha\}.$$

VaR-luvun voi tulkita olevan suurin mahdollinen tappio jota ei ylitetä todennäköisyydellä α (McNeil ym., 2005). Matemaattisesti VaR-luvussa etsitään pienin piste t johon mennessä pätee $\mathbb{P}(X \leq t) \geq \alpha$, jolloin käytännössä kyseessä on satunnaismuuttujan X α -kvantiilin selvittäminen. Esimerkiksi tappiojakaumalle X luottamustasolla $\alpha = 0,95$ laskettu VaR-luku

$$VaR_{0,95}(X) = \inf\{t \in \mathbb{R} \mid \mathbb{P}(X \leq t) \geq 0,95\} = 250\,000 \text{ €}$$

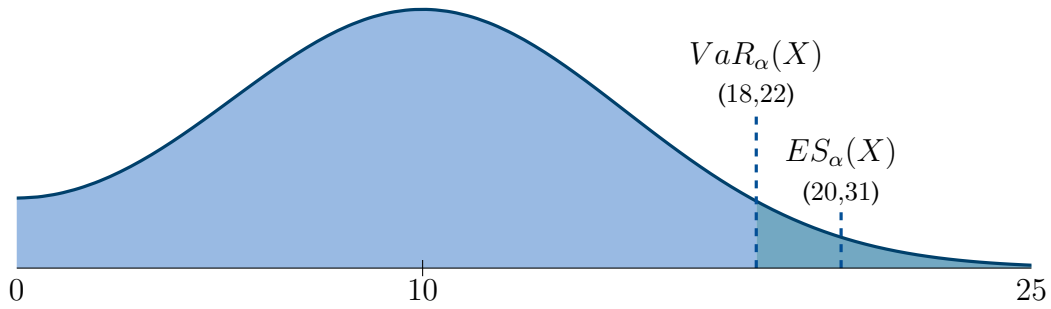
tarkoittaa sitä, että 95 prosentin todennäköisyydellä tappion suuruus on pienempää kuin 250 000 € ja viiden prosentin todennäköisyydellä tappion suuruus on vähintään 250 000 €.

Vaikka rajapyykkietieto on tärkeä suuntaa antava tieto tappiojakaumasta, niin joskus voi kuitenkin olla tarpeen saada lisätietoa tämän tason ylittävistä tapahtumista. Todennäköisyysmassa voi olla epätasaisesti jakautunut rajapyykin jälkeen, jolloin mahdolliset tappiot voivat olla luottamustason $\alpha = 0,95$ jälkeen hyvinkin suuria. Edellisen esimerkin tilanteessa VaR-luku ei kerro meille onko kyseisen luottamustason ylittävän tappion suuruus luultavammin 275 000 € tai 750 000 €. Tätä puutetta kutsutaan kirjallisuudessa häntäriskiksi (*Tail risk*), ja se voi johtaa tilanteisiin missä sijoituspäätöksiä tehdessä paksuhäntäinen tappiojakauma nähdään parempana vaihtoehtona kuin kevythäntäinen tappiojakauma (Yamai & Yoshida, 2005). Tämän puutteen korjaamiseksi voidaan VaR-luvun lisäksi käyttää odotettua tappiota, joka keskiylitysfunktion tavoin tarkastelee odotusarvotasolla tappiojakaumaa jonkin pisteen jälkeen.

Määritelmä 4.18. (Odotettu tappio) Olkoon X tappioita kuvastava satunnaismuuttuja. Luottamustason $\alpha \in (0, 1)$ ylittävä odotettu tappio ES_α (*Expected shortfall*) lasketaan kaavalla

$$ES_\alpha(X) = \mathbb{E}(X \mid X > VaR_\alpha(X)) = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_\gamma(X) \, d\gamma.$$

Täten odotettu tappio kertoo ehdollisen odotusarvon kun satunnaismuuttuja X on ylittänyt tason $VaR_\alpha(X)$ (Yamai & Yoshida, 2005). Koska $VaR_\alpha(X)$ on reaaliluku, niin tällöin sen voi yksinkertaistetusti tulkita olevan samankaltainen kynnys kuin edellisissä luvuissa esitetty ylitetaso u . Piirtämällä molemmat riskimitat taitetun normaalijakauman tiheysfunktion kanssa samaan kuvaan kiinnitetyllä luotettavuustasolla on mahdollista nähdä paremmin mitä kumpikin riskimita kuvastaa.



Kuva 4.2: Taitetun normaalijakauman $FN(10, 5)$ tiheysfunktio ja tasolla $\alpha = 0,95$ laskettu $VaR_\alpha(X)$ -luku $(18,22)$ ja odotettu tappio $ES_\alpha(X)$ $(20,31)$.

Kuvasta 4.2 voidaan nähdä, että odotettu tappio kertoo satunnaismuuttujan $VaR_\alpha(X)$ -luvun ylittävän osan odotusarvon. Tämä muistuttaa keskiylitysfunktion toimintaa, ja huomataankin että odotetusta tappiosta voidaan myös helposti muokata keskiylitysfunktio. Tähän riittää pelkästään vakion u vähentäminen odotetusta tappiosta, jolloin saadaan

$$\underbrace{\mathbb{E}(X \mid X > u) - u}_{\text{Odotettu tappio}} = \mathbb{E}(X \mid X > u) - \mathbb{E}(u \mid X > u) = \underbrace{\mathbb{E}(X - u \mid X > u)}_{\text{Keskiylitysfunktio}}.$$

Täten keskiylitysfunktiosta on mahdollista saada laskettua odotettu tappio lisäämällä luku u keskiylitysfunktion $e(u)$ arvoon pisteessä u . Merkitsemällä tällä tavalla laskettua odotettua tappiota $v(u) = e(u) + u$ voimme huomata, että proposition 4.8 kolmannen kohdan perusteella odotettu tappio on kasvava funktio. Kasvavuus ei kuitenkaan ole keskiylitysfunktion ominaisuus, jonka vuoksi vaikuttaisi siltä että keskiylitysfunktio ja odotettu tappio havainnollistavat satunnaismuuttujaa X eri tavoilla. Luvussa 5 tarkastellaan tarkemmin sitä mitä keskiylitysfunktion kasvavuus kertoo itse satunnaismuuttujasta X .

Vaikka VaR-lukua käyttäessä altistuu häntäriskille, niin sen korvaaminen odotetulla tappiolla ei kuitenkaan tee muodostettavista riskiarvioista automaattisesti parempia. Kirjallisuudessa on osoitettu, että kun satunnaismuuttuja X on paksuhäntäinen jakauma, niin tällöin odotetun tappion antamat arviot tappion suuruudesta voivat heitellä enemmän kuin VaR-luvun antama arvio (Yamai & Yoshida, 2005). VaR-luku on myös yksi yleisimmistä riskienhallinnassa käytetyistä riskimitoista ja samalla se on oleellinen osa finanssialaa koskevassa säätelystä (McNeil ym., 2005). Käyttämällä VaR-lukua ja odotettua tappiota samanaikaisesti saadaan muodostettua monipuolisempi katsaus satunnaismuuttujan X aiheuttamasta riskistä kuin rajoittamalla pelkästään yhden riskimitan käyttöön (McNeil ym., 2005; Yamai & Yoshida, 2005).

Riskimitoista sekä niiden käytöstä finanssi- ja vakuutuslalla on mahdollista lukea enemmän lähteissä (McNeil ym., 2005, luku 2.2) ja (Asmussen & Steffensen, 2020, luku X).

5 | Paksuhäntäinen keskiylitys

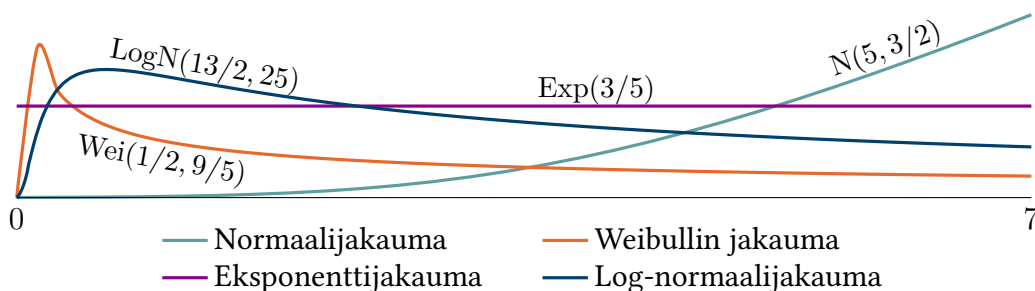
Kuvasta 4.1 nähdään, että luvussa 4 lasketuista keskiylitysfunktioista pelkästään paksuhäntäisen yleistetyn Pareto-jakauman keskiylitysfunktio on kasvava. Muiden jakaumien keskiylitysfunktioissa vastaavaa kasvavaa käyttäytymistä ei ole nähtävissä. Mitkä tekijät voisivat aiheuttaa tällaista käyttäytymistä? Tässä luvussa tarkastellaankin tarkemmin paksuhäntäisyyttä ja sen vaikutusta keskiylitysfunktion käyttäytymiseen. Ennen tätä on hyvä kuitenkin havainnollistaa kuinka "vaarallisia" eri jakaumat ovat. Yksi jakaumien vaarallisuuden havainnollistamiseen soveltuva työkalu on *riskitiheys* (Kleiber & Kotz, 2003).

Määritelmä 5.1. (Riskitiheys) Kertymäfunktion $F(x)$ määräämän satunnaismuuttujan X riskitiheys (*Hazard rate*) $\lambda(x)$ on muotoa

$$\lambda(x) = \frac{f(x)}{1 - F(x)} = -\frac{d}{dx} \ln \bar{F}(x), \quad x \geq 0,$$

missä $f(x)$ on satunnaismuuttujan X tiheysfunktio. Riskitiheydestä käytetään myös nimiä *hasardifunktio* luotettavuusteoriassa ja *kuolevuus* henkivakuutusmatematiikassa.

Riskitiheys $\lambda(x)$ saa pelkästään epänegatiivisia arvoja, jotka kuvastavat kuinka nopeasti suurten tapahtumien riski vähenee pisteessä x . Mitä suuremman arvon riskitiheys saa pisteessä x , niin sitä pienempi suuren tapahtuman aiheuttama riski on. Tällöin jos riskitiheys on matala, suurten tapahtumien riski vähenee hitaasti kyseisessä pisteessä. Täten matalan riskitiheyden suurilla x omaavat jakaumat ovat vaarallisempia jakaumia kuin korkean riskitiheyden omaavat jakaumat. (Kleiber & Kotz, 2003)



Kuva 5.1: Normaalijakauman, eksponenttijakauman, Weibullin- ja log-normaalijakauman riskitiheydet välillä $[0, 7]$. Jakaumien parametrit ovat valittu riskitiheyden käyttäytymisen havainnollistamisen tueksi.

Kuvan 5.1 mukaan paksuhäntäisten Weibullin ja log-normaalijakauman riskitiheydet ovat laskevia. Tällöin suurilla x ei ole perustetta olla odottamatta suuria mahdollisia tapahtumia, vaan x kasvaessa suurien tapahtumien riski kasvaa. Vastaavasti kevythäntäisen nor-

maalijakauman riskitiheys on kasvava, jolloin suurien tapahtumien riski pienenee jatkuvasti. Muistinmenetyksomaisuus näkyy eksponenttijakauman riskitiheydessä, jolloin suurien tapahtumien riski on yhtä suuri riippumatta x arvosta. Täten suurten tapahtumien riskien näkökulmasta ei ole yhdentekevää mitä jakaumaa käyttää häntätapahtumien mallintamiseen.

Käytännön sovelluskohteissa ollaankin usein kiinnostuneita äärimmäisten ilmiöiden vakavuudesta. Varsinkin vakuutus- ja finanssialalla äärimmäisten ilmiöiden perusteellinen mallinnus on oleellista yrityksen vakavaraisuuden säilyttämisen vuoksi (McNeil ym., 2005). Paksuhäntäiset satunnaismuuttajat tarjoavat hyvän ehdokkaan tällaisten tapahtumien mallinnukseen. Varsinkin paksuhäntäisiin jakaumiin lukeutuva *alieksponentiaalisten jakaumien* joukon on todettu kuvaavan hyvin oikean maailman harvinaisia ilmiöitä. (Embrechts ym., 1997)

Tämän vuoksi luvussa 5.1 esitellään tarkemmin paksuhäntäisyyden määritelmä, sen alaluokat ja miten keskiylitysfunktiot käyttäytyvät paksuhäntäisillä jakaumilla. Luvussa 5.2 esitellään tulos, jota hyödyntämällä keskiylitysfunktioita voi hyödyntää päätelyssä pelkästään asymptotiikkaan nojaten. Lopulta luvussa 5.3 muodostetaan yhteenveto miten jakauman hännän tyyppin voi päätellä keskiylitysfunktioita käyttämällä.

5.1 | Alieksponentiaaliset jakaumat

Työn alussa esitetyssä määritelmässä (2.11) jakauman sanotaan olevan (oikealle) paksuhäntäinen jos momentit generoiva funktio $M_X(t)$ on ääretön kaikilla $t > 0$. Tästä määritelmästä ei suoraan saada arvioita häntäfunktion tyypistä, sillä pelkästään joukkoon

$$\mathcal{K} = \{X \mid M_X(t) = \infty \text{ kaikilla } t > 0\}$$

kuuluminen ei anna tarkentavaa lisätietoa häntäfunktion käyttäytymisestä. Joukko \mathcal{K} on laaja, ja se pitää sisällään monenlaisia paksuhäntäisiä satunnaismuuttujia. Mitä kaikkea tästä määritelmästä voi päätellä? Jos jakauma on kevythäntäinen, niin tällöin lemmän 3.3 mukaan häntäjakauman suppenemisnopeudella saadaan eksponentiaalinen yläraja jollain $t > 0$. Jos jakauma on paksuhäntäinen, niin tällaista ylärajaa ei ole olemassa millään $t > 0$. Täten paksuhäntäisillä jakaumilla häntäfunktion suppenemisnopeudelle ei ole olemassa eksponentiaalista ylärajaa.

Tämä tieto häntäjakauman suppenemisnopeuden ylärajan puuttumisesta on vieläkin varsin yleinen jokaiselle joukon \mathcal{K} alkiolle. Vielä tarkempi luokittelu on mahdollista saada eritelemällä jakaumia häntäfunktioiden käyttäytymisen perusteella. Seuraavassa määritelmässä esitellään kirjallisuudessa yleisessä käytössä oleva paksuhäntäisten satunnaismuuttujien luokittelu.

Määritelmä 5.2. (Paksuhäntäisyyden luokat) Paksuhäntäiset jakaumat voidaan luokitella kolmeen pääluokkaan:

- *Paksuhäntäinen jakauma:* Satunnaismuuttujan X sanotaan olevan (oikealle) paksuhäntäinen jos momentit generoiva funktio $M_X(t)$ on ääretön kaikilla $t > 0$. Paksuhäntäisten jakaumien joukkoa merkitään \mathcal{K} .

- *Pitkähäntäinen jakauma*: Satunnaismuuttujan X jakauman sanotaan olevan (oikealle) pitkähäntäinen jos ja vain jos kaikilla $t > 0$ pätee

$$\lim_{x \rightarrow \infty} \mathbb{P}(X > x + t \mid X > x) = 1.$$

Pitkähäntäisten jakaumien joukkoa merkitään \mathcal{L} .

- *Alieksponentiaalinen jakauma*: Olkoot X, X_1, \dots, X_n riippumattomia ja samoin jakautuneita satunnaismuuttujia ja $S_n = X_1 + \dots + X_n$. Satunnaismuuttujan X jakauman sanotaan olevan alieksponentiaalinen jos ja vain jos häntäfunktiolle pätee

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}(S_n > x)}{\mathbb{P}(X > x)} = n$$

kaikilla $n \geq 2$. Alieksponentiaalisten jakaumien joukkoa merkitään \mathcal{S} .

Jokainen määritelmässä 5.2 esitetty väite asettaa häntäfunktion käyttäytymiselle lisäehtoja. Täten jokainen alieksponentiaalinen jakauma on myös pitkähäntäinen jakauma. Samoin jokainen pitkähäntäinen jakauma on myöskin paksuhäntäinen. On kuitenkin huomiotava että $\mathcal{L} \not\subset \mathcal{S}$ ja $\mathcal{K} \not\subset \mathcal{L}$. (Embrechts ym., 1997)

Näistä kolmesta luokasta alieksponentiaalisten satunnaismuuttujien luokka on sovelluskohteita silmälläpitäen yleisin. Vakuutusmatemaattisissa sovelluksissa kaikki yleisessä käytössä olevat paksuhäntäiset satunnaismuuttujat ovat alieksponentiaalisia. Alieksponentiaalisilla satunnaismuuttujilla useimmiten tarkoitetaan positiivisia satunnaismuuttujia, joka tekee niistä sopivia vakuutusmatemaattisiin sovelluksiin. (Embrechts ym., 1997)

Määritelmä 5.3. (Asymptoottinen yhtäsuuruus) Funktioita $f(x)$ ja $g(x)$ sanotaan asymptoottisesti yhtäsuuriksi, jos osamäärän $f(x)/g(x)$ raja-arvolle pätee

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 1.$$

Tällöin merkitään $f(x) \sim g(x)$, jolla tarkoitetaan funktioiden käyttäytymistä kun $x \rightarrow \infty$.

Asymptoottiset arviot funktioiden yhtäsuuruuksista ovat vakuutus- ja finanssivakuutussovelluksissa tärkeitä (Embrechts ym., 1997). Varsinkin häntäjakaumien arvioinnissa asymptoottisen yhtäsuuruuden tietäminen kertymäfunktioiden välillä mahdollistaa monipuolisempaa päättelyä. Esimerkiksi tieto $\bar{F}(x) \sim \bar{G}(x)$ kertoo, että tarpeeksi suurilla x häntäfunktiot ovat melkein yhtäsuuret. Tällöin laskennassa on mahdollista käyttää pelkästään toista näistä kahdesta häntäfunktiosta.

Yksi tapa käsitellä kyseisiä satunnaismuuttujia on tarkastella riippumattomien ja samoin jakautuneiden satunnaismuuttujien X_1, \dots, X_n summan

$$\mathbb{P}(X_1 + \dots + X_n < x) = \mathbb{P}(S_n < x)$$

käyttäytymistä suurilla x . Tällaisen summan jakauman arviointi voi olla mielekästä esimerkiksi vakuutuskontekstissa, missä jokainen X_i ($i \in \{1, \dots, n\}$) kuvastaa yksittäisen vahingon suuruutta ja S_n vakuutusyhtiön korvattavaa kokonaisvahinkomäärää.

Varsinkin paksuhäntäisten satunnaismuuttujien tilanteessa summan S_n jakauman asymptoottinen käyttäytyminen voi aiheuttaa ongelmia vakavaraisuuslaskennassa. Miten yksittäinen todella suuri tapahtuma X vaikuttaa summajakauman käyttäytymiseen? Koska alieksponentiaaliset satunnaismuuttujat ovat yleisiä sovelluskohteissa, niin täten on mahdollista käyttää seuraavaan proposition tulosta summan häntäfunktion arvojen laskemiseen suurilla arvoilla.

Propositio 5.4. *Alieksponentiaalisilla satunnaismuuttujilla summan ja maksimin häntäntä-funktioille pätee*

$$\mathbb{P}(S_n > x) \sim \mathbb{P}(M_n > x)$$

kaikilla $n \geq 2$ kun $x \rightarrow \infty$.

Todistus. Olkoot X, X_1, \dots, X_n riippumattomia ja samoin jakautuneita alieksponentiaalisia satunnaismuuttujia. Maksimin M_n häntäfunktion asympotoottista käyttäytymistä määrätessä voidaan tarkastella raja-arvoa

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}(M_n > x)}{n\mathbb{P}(X > x)} = \frac{1}{n} \lim_{x \rightarrow \infty} \frac{1 - F^n(x)}{1 - F(x)} = \frac{1}{n} \lim_{x \rightarrow \infty} \sum_{k=0}^{n-1} F^k(x)$$

geometrisen sarjan esitysmuotoa hyödyntämällä. Koska $F(x) \in [0, 1]$ kaikilla $x \in \mathbb{R}$, niin täten summa on äärellinen ja raja-arvo voidaan viedä summan sisään. Tällöin saadaan

$$\frac{1}{n} \sum_{k=0}^{n-1} \lim_{x \rightarrow \infty} F^k(x) = \frac{1}{n} \sum_{k=0}^{n-1} 1 = \frac{n}{n} = 1,$$

josta seuraa

$$\mathbb{P}(M_n > x) \sim n\mathbb{P}(X > x).$$

Summan S_n häntäfunktion asympotoottisessa arvioinnissa voidaan käyttää hyväksi alieksponentiaalisuuden määritelmää, josta saadaan

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}(S_n > x)}{\mathbb{P}(X > x)} = n \Leftrightarrow \lim_{x \rightarrow \infty} \frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(X > x)} = 1 \Leftrightarrow \mathbb{P}(S_n > x) \sim n\mathbb{P}(X > x).$$

Täten tarpeeksi suurilla x pätee

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}(S_n > x)}{\mathbb{P}(M_n > x)} = \lim_{x \rightarrow \infty} \frac{n\mathbb{P}(X > x)}{n\mathbb{P}(X > x)} = 1,$$

joka todistaa väitteen. □

Toisin sanoen alieksponentiaalisilla satunnaismuuttujilla summan S_n jakauman määrää summan suurin termi $\max\{X_n\}$. Tämän voi tulkita kuvastavan tilannetta, missä yksi erittäin suuri havainto määrittää koko summan S_n tason. Tällöin muiden havaintojen aiheuttama vaihtelu summassa ei enää vaikuta saataviin todennäköisyyksiin merkittävästi, ja laskennassa voidaan käyttää pelkästään maksimin jakaumaa. (Embrechts ym., 1997)

On myös hyvä huomauttaa, että propositiossa 5.4 esitetty väite häntäfunktioiden ekvivalenssista vaatii riippumattomuusoletuksen. Vakuutusmatemaattisissa sovelluksissa riippumattomuusoletuksen käyttö on usein perusteltua, mutta finanssimatemaattisissa sovelluksissa riippumattomuusoletuksen käyttö esimerkiksi osakekurssin kehitystä mallintaessa voi olla kyseenalaista. Osakkeiden hintojen suuret muutokset ovat historiallisesti olleet ryhmittyneitä. Yhtenä esimerkkinä tästä osakehintojen kehityksen ryhmittymisestä on keväällä 2020 tapahtunut koronaviruspandemian aiheuttanut pörssiromahdus. Tämän vuoksi finanssimatemaattisissa sovelluksissa ei voida suoraan käyttää samoja tuloksia kuin vakuutusmatemaattisissa sovelluksissa. (Embrechts ym., 1997)

Alieksponentiaalisten satunnaismuuttujien joukkoon kuuluvista satunnaismuuttujista useimmiten käytetään jakaumaa joka kuuluu yhteen seuraavista jakaumatyypeistä: *säännöllisesti vaihtelevat*, (paksuhäntäisen) *Weibullin tyyppiset* ja *lognormaalit* (Asmussen & Steffensen, 2020). Esimerkeissä 5.5–5.8 lasketaan edellä mainituille kolmelle alieksponentiaalisel-

le jakaumalle asymptoottisesti ekvivalentit keskiylitysfunktiot, jolloin saamme myös tietää miten jakaumia vastaavat keskiylitysfunktiot käyttäytyvät suurilla arvoilla.

Esimerkki 5.5. Olkoon häntäfunktio $\bar{F}(x) = L(x)/x^\alpha$, missä $\alpha > 1$ ja $L(x)$ on määritelmän 2.12 mukainen hitaasti vaihteleva funktio. Tällöin $\bar{F}(x)$ on säännöllisesti vaihtelevan satunnaismuuttujan häntäfunktio. Osamäärää

$$\frac{e(x)}{x/(\alpha-1)} = \frac{\frac{x^\alpha}{L(x)} \int_x^\infty L(t)/t^\alpha dt}{x/(\alpha-1)} = \frac{(\alpha-1)x^\alpha \int_x^\infty t^{-\alpha} L(t) dt}{xL(x)}$$

tarkastelemalla on mahdollista määrätä keskiylitysfunktion $e(x)$ käyttäytyminen kun $x \rightarrow \infty$. Käyttämällä osoittajan integraaliin *Karamatan lausetta* (Lause A.9) suurilla x saadaan

$$\int_x^\infty t^{-\alpha} L(t) dt \sim -(-\alpha+1)^{-1} x^{-\alpha+1} L(x).$$

Täten rajankäynnissä $x \rightarrow \infty$ alkuperäinen osamäärä saa arvokseen

$$\frac{(\alpha-1)x^\alpha \int_x^\infty L(t)/t^\alpha dt}{xL(x)} = \frac{(\alpha-1)x^\alpha x^{-\alpha+1} L(x)}{x(\alpha-1)L(x)} = 1,$$

josta saamme pääteltyä

$$e(x) \sim \frac{x}{\alpha-1}.$$

Esimerkki 5.6. Olkoon häntäfunktio $\bar{F}(x) = \exp\{-\alpha x^\beta\}$ Weibullin jakauman häntäfunktio parametrein $\alpha = 1$ ja $0 < \beta < 1$. Tällöin keskiylitysfunktion asymptoottinen käyttäytyminen saadaan selvitettyä osamäärän

$$\frac{e(x)}{x^{1-\beta}/\beta} = \frac{\frac{1}{\exp\{-x^\beta\}} \int_x^\infty \exp\{-t^\beta\} dt}{x^{1-\beta}/\beta} = \beta \frac{\int_x^\infty \exp\{-t^\beta\} dt}{\exp\{-x^\beta\} x^{1-\beta}}$$

rajakäyttäytymisen avulla. Koska haluamme näyttää rajakäyttäytymiselle ekvivalenssin, niin tällöin yllä olevan osamäärän tulisi saada rajalla $x \rightarrow \infty$ arvokseen yksi. Koska sekä osoittaja että nimittäjä suppenevat kohti nollaa, niin täten raja-arvon selvittämisessä voidaan käyttää L'Hôpitalin sääntöä.

Valinnalla $u \in (x, \infty)$ osoittajassa oleva integraali voidaan hajoittaa osiin

$$\int_x^\infty \exp\{-t^\beta\} dt = \int_x^u \exp\{-t^\beta\} dt + \int_u^\infty \exp\{-t^\beta\} dt,$$

jonka perusteella osoittajan derivaatta voidaan ilmaista muodossa

$$\frac{d}{dx} \int_x^\infty \exp\{-t^\beta\} dt = \frac{d}{dx} \int_x^u \exp\{-t^\beta\} dt + \frac{d}{dx} \int_u^\infty \exp\{-t^\beta\} dt. \quad (5.7)$$

Koska Weibullin jakauma on epänegatiivinen ja sillä on äärellinen odotusarvo, niin täten yhtälön (4.11) kaltaisesta epänegatiivisten satunnaismuuttujien odotusarvon kaavaa hyödyntämällä integraalille saadaan äärellinen yläraja

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t) dt \geq \int_u^\infty \mathbb{P}(X \geq t) dt.$$

Täten yhtälön (5.7) jälkimmäinen termi on äärellinen vakio, jonka derivaatta muuttujan x suhteen on nolla. Saman yhtälön ensimmäinen termi saadaan Leibnizin integroimissääntöä

(Lemma A.10) käyttämällä muotoon

$$\begin{aligned} \frac{d}{dx} \int_x^u \exp\{-t^\beta\} dt &= \exp\{-u^\beta\} \frac{d}{dx} u - \exp\{-x^\beta\} \frac{d}{dx} x + \int_x^u \frac{d}{dx} \exp\{-t^\beta\} dt \\ &= -\exp\{-x^\beta\}, \end{aligned}$$

jolloin saadaan

$$\frac{d}{dx} \int_x^\infty \exp\{-t^\beta\} dt = -\exp\{-x^\beta\}.$$

Nimittäjästä saadaan tulon derivaatalla

$$\frac{d}{dx} (\exp\{-x^\beta\} x^{1-\beta}) = -\beta x^{\beta-1} \exp\{-x^\beta\} x^{1-\beta} + (1-\beta) x^{-\beta} \exp\{-x^\beta\}.$$

Tällöin L'Hôpitalin säännön mukaisesta osamäärästä tulee

$$\begin{aligned} \beta \frac{\frac{d}{dx} \int_x^\infty \exp\{-t^\beta\} dt}{\frac{d}{dx} (\exp\{-x^\beta\} x^{1-\beta})} &= \beta \frac{-\exp\{-x^\beta\}}{-\beta x^{\beta-1} x^{1-\beta} \exp\{-x^\beta\} + (1-\beta) x^{-\beta} \exp\{-x^\beta\}} \\ &= \frac{-\beta}{-\beta + (1-\beta) x^{-\beta}}, \end{aligned}$$

josta ottamalla raja-arvo $x \rightarrow \infty$ saadaan

$$\frac{-\beta}{-\beta + (1-\beta) x^{-\beta}} \xrightarrow{x \rightarrow \infty} \frac{-\beta}{-\beta} = 1.$$

Täten L'Hôpitalin säännön perusteella suurilla x pätee

$$e(x) \sim \frac{x^{1-\beta}}{\beta}.$$

Esimerkki 5.8. Olkoon häntäfunktio $\bar{F}(x)$ lognormaalijakauman häntäfunktio, jota suurilla x voidaan approksimoida arviolla $\bar{F}(x) \sim \frac{\sigma}{\sqrt{2\pi} \log x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}$. Osamäärää

$$\frac{e(x)}{\sigma^2 x / \log x} = \frac{1}{\sigma^2} \frac{\int_x^\infty \frac{\exp\{-(\log t - \mu)/2\sigma^2\}}{\log t} dt}{\frac{x}{\log^2 x} \exp\{-(\log x - \mu)^2/2\sigma^2\}}$$

tarkastelemalla on mahdollista näyttää keskiylitysfunktion $e(x)$ asymptoottinen käyttäytyminen jos osamäärä suppenee kohti lukua yksi kun $x \rightarrow \infty$. Koska sekä osoittaja että nimittäjä suppenevat kohti nollaa kun $x \rightarrow \infty$, niin tällöin raja-arvon selvittämisessä voimme käyttää L'Hôpitalin sääntöä. Koska lognormaalin satunnaismuuttujan odotusarvo on äärellinen, niin täten esimerkin 5.6 mukaisesti Leibnizin integroimissääntöä käyttämällä osoittaja saadaan muotoon

$$\frac{d}{dx} \int_x^\infty \frac{\exp\{-(\log t - \mu)/2\sigma^2\}}{\log t} dt = -\frac{\exp\{-(\log x - \mu)/2\sigma^2\}}{\log x}.$$

Nimittäjä saadaan osamäärän derivoimissääntöä käyttämällä muotoon

$$\begin{aligned} &\frac{d}{dx} \frac{x \exp\{-(\log x - \mu)^2/2\sigma^2\}}{\log^2 x} \\ &= \exp\{-(\log x - \mu)^2/2\sigma^2\} \left(\frac{\mu - \log x}{\sigma^2 \log^2 x} + \frac{\log x - 2}{\log^3 x} \right). \end{aligned}$$

Sijoittamalla saadut derivaatat L'Hôpitalin säännön mukaiseen osamäärään saadaan

$$\begin{aligned} & \frac{\frac{d}{dx}e(x)}{\frac{d}{dx}[\sigma^2 x / \log x]} \\ &= \frac{1}{\sigma^2} \frac{-\exp\{-(\log x - \mu)^2 / 2\sigma^2\} / \log x}{\exp\{-(\log x - \mu)^2 / 2\sigma^2\} \left(\frac{\mu - \log x}{\sigma^2 \log^2 x} + \frac{\log x - 2}{\log^3 x} \right)} \\ &= \frac{1}{\sigma^2} \left(\frac{\log x - \mu}{\sigma^2 \log x} - \frac{\log x - 2}{\log^3 x} \right)^{-1} \\ &= \left(1 - \frac{\mu}{\log x} - \frac{\sigma^2}{\log^2 x} + \frac{2\sigma^2}{\log^3 x} \right)^{-1}. \end{aligned}$$

Ottamalla raja-arvo $x \rightarrow \infty$ edellisestä osamäärästä saadaan

$$\left(1 - \frac{\mu}{\log x} - \frac{\sigma^2}{\log^2 x} + \frac{2\sigma^2}{\log^3 x} \right)^{-1} \xrightarrow{x \rightarrow \infty} 1.$$

Täten suurilla x pätee

$$e(x) \sim \frac{\sigma^2 x}{\log x}.$$

Näin ollen on saatu näytettyä, että suurilla x kolmella edellä mainituilla alieksponentiaalisella satunnaismuuttujalla keskiylitysfunktio käyttäytyy

$$e(x) \sim \frac{x}{\alpha - 1}, \quad e(x) \sim \frac{x^{1-\beta}}{\beta} \quad \text{tai} \quad e(x) \sim \frac{\sigma^2 x}{\log x},$$

missä $\alpha > 1, \beta \in (0, 1)$ ja $\sigma > 0$. Koska nämä esitysmuodot on saatu laskettua, heräänkin kysymys, voidaanko näitä keskiylitysfunktioita käyttää esityslauseen (Lause 4.14) kaltaisesti suurilla x hyödyksi jakauman määrittämisessä? Jos tämä olisi mahdollista, niin tällöin jakauman hänälle voitaisiin saada määriteltyä sitä vastaava jakauma keskiylitysfunktiota hyödyntäen. Tähän asiaan paneudutaankin tarkemmin seuraavassa luvussa.

Alieksponentiaalisista satunnaismuuttujista on mahdollista lukea kattavempi tarkastelu lähteistä (Embrechts ym., 1997) ja (Peters & Shevchenko, 2015).

5.2 | Asymptoottinen jakauma

Esimerkeissä 4.5 ja 4.6 saatiin laskettua keskiylitysfunktiolle suljettu esitysmuoto käyttämällä tunnettua häntäfunktiota. Näiden keskiylitysfunktioiden avulla voidaan esityslauseetta (Lause 4.14) käyttämällä määritellä satunnaismuuttujan jakauma positiivisilla reaaliluvuilla, joka tosin vaatii myös keskiylitysfunktion analyttisen esitysmuodon tuntemisen kaikilla $x \in [0, \infty)$.

Esimerkeissä 5.5–5.8 saatiin tunnetuille häntäfunktioille näytettyä keskiylitysfunktion asymptoottinen käyttäytyminen tarpeeksi suurilla x . Tällöin esityslauseen käyttö ei ole perusteltua, sillä tiedämme häntäfunktiosta pelkästään asymptoottisen käyttäytymisen, emmekä käyttäytymistä pienillä x . Asymptoottinen käyttäytyminen kuitenkin kertoo jotain taustalla olevan jakauman käyttäytymisestä. Artikkelissa (Hall & Wellner, 1981) esitetyn proposition mukaan jakauman päättely on mahdollista myös tällaisissa tilanteissa. Seuraa-

va propositio ja sen todistus on esitetty mukailien kyseisessä artikkelissa olevaa propositiota ja sen todistusta.

Propositio 5.9. Oletetaan että jakaumien F ja G odotusarvot μ_F ja μ_G ovat äärellisiä ja että kyseisten jakaumien keskiylitysfunktioille $e_F(x)$ ja $e_G(x)$ pätee

$$(i) e_F(x) \sim e_G(x) \text{ ja } (ii) -\log(e_G(x)\overline{G}(x)) = O(-\log\overline{G}(x)) \text{ kun } x \rightarrow \infty.$$

Tällöin $-\log\overline{F}(x) \sim -\log\overline{G}(x)$ tai yhtäpitävästi $\overline{F}(x) = \overline{G}(x)^{1+o(1)}$.

Todistus. Merkitään $r(x) = \overline{F}(x)/\overline{G}(x)$. Häntäfunktion esitysmuotoa (4.15) hyödyntämällä funktio $r(x)$ voidaan ilmaista muodossa

$$\begin{aligned} r(x) &= \frac{\overline{F}(x)}{\overline{G}(x)} = \frac{\frac{e_F(0)}{e_F(x)} \exp\left\{-\int_0^x \frac{1}{e_F(t)} dt\right\}}{\frac{e_G(0)}{e_G(x)} \exp\left\{-\int_0^x \frac{1}{e_G(t)} dt\right\}} \\ &= \frac{\mu_F e_G(x)}{\mu_G e_F(x)} \exp\left\{\int_0^x \frac{1}{e_G(t)} - \frac{1}{e_F(t)} dt\right\} \\ &= \frac{\mu_F e_G(x)}{\mu_G e_F(x)} \exp\left\{-\int_0^x \frac{h(t)}{e_G(t)} dt\right\}, \end{aligned} \quad (5.10)$$

missä funktio $h(x)$ on

$$h(x) = \frac{e_G(x)}{e_F(x)} - 1.$$

Ottamalla logaritmi funktiosta $r(x)$ saadaan yhtälö

$$\begin{aligned} \log r(x) &= \log\left(\frac{\overline{F}(x)}{\overline{G}(x)}\right) \\ &= \log\left(\frac{\mu_F}{\mu_G}\right) + \log\left(\frac{e_G(x)}{e_F(x)}\right) - \int_0^x \frac{h(t)}{e_G(t)} dt, \end{aligned} \quad (5.11)$$

missä esiintyvälle integraalille on mahdollista kehittää vaihtoehtoinen esitysmuoto. Arvioimalla kyseistä integraalia ylöspäin saadaan

$$\begin{aligned} \int_0^x \frac{h(t)}{e_G(t)} dt &\leq \left| \int_0^x \frac{h(t)}{e_G(t)} dt \right| \\ &= \left| \int_0^s \frac{h(t)}{e_G(t)} dt \right| + \left| \int_s^x \frac{h(t)}{e_G(t)} dt \right| \\ &\leq \left| \int_0^s \frac{h(t)}{e_G(t)} dt \right| + \int_s^x \frac{|h(t)|}{e_G(t)} dt. \end{aligned} \quad (5.12)$$

Koska oletuksen (i) mukaan pätee $h(t) \rightarrow 0$ kun $t \rightarrow \infty$, niin tällöin raja-arvon ominaisuuksien perusteella arvio $|h(t)| < \delta$ pätee kaikilla $\delta > 0$ kun $t > s = s_\delta$. Tällöin yhtälölle

(5.12) saadaan arvioitua yläraja

$$\begin{aligned}
\left| \int_0^s \frac{h(t)}{e_G(t)} dt \right| + \int_s^x \frac{|h(t)|}{e_G(t)} dt &\leq \left| \int_0^s \frac{h(t)}{e_G(t)} dt \right| + \delta \int_s^x \frac{1}{e_G(t)} dt \\
&= \underbrace{\left| \int_0^s \frac{h(t)}{e_G(t)} dt \right| - \delta \int_0^s \frac{1}{e_G(t)} dt}_{=A(s)} + \delta \int_0^x \frac{1}{e_G(t)} dt \\
&= A(s) + \delta \int_0^x \frac{1}{e_G(t)} dt.
\end{aligned} \tag{5.13}$$

Yhtäsuuruutta (4.12) käyttämällä edellisen yhtälön integraali voidaan esittää muodossa

$$\begin{aligned}
\delta \int_0^x \frac{1}{e_G(t)} dt &= -\delta \log \left(\frac{e_G(x) \overline{G}(x)}{\mu_G} \right) \\
&= -\delta \log(e_G(x) \overline{G}(x)) + \delta \log \mu_G.
\end{aligned} \tag{5.14}$$

Oletuksen (ii) mukaan pätee $-\log(e_G(x) \overline{G}(x)) = O(-\log \overline{G}(x))$, missä ordo-notaatio O tarkoittaa määritelmän A.8 mukaista joukkoa

$$O(-\log \overline{G}(x)) = \{f(x) \mid \exists c > 0 \exists x_0 > 0 \forall x \geq x_0 : 0 \leq f(x) \leq -c \log \overline{G}(x)\}.$$

Kyseistä joukkoa hyödyntämällä saadaan yhtälössä (5.14) olevalle termille $-\log(e_G(x) \overline{G}(x))$ yläraja

$$-\delta \log(e_G(x) \overline{G}(x)) \leq \delta c \log \overline{G}(x),$$

missä $c > 0$ on jokin joukon $O(-\log \overline{G}(x))$ vaatima positiivinen kerroin. Valitaan $c' = \delta c$. Koska epäyhtälössä (5.13) esitetty yläraja pätee raja-arvon ominaisuuksien perusteella kaikilla $\delta > 0$, niin tällöin myös termi c' voi saada minkä tahansa positiivisen arvon riippumatta termin $c > 0$ arvosta. Tällöin voimme o -notaation mukaisesti merkitä

$$o(-\log \overline{G}(x)) = \{f(x) \mid \forall c' > 0 \exists x_0 > 0 \forall x \geq x_0 : 0 \leq f(x) \leq -c' \log \overline{G}(x)\} \tag{5.15}$$

Hyödyntämällä joukkoa $o(-\log \overline{G}(x))$ yhtälö (5.14) saadaan ilmaistua muodossa

$$\delta \int_0^x \frac{1}{e_G(t)} dt = o(-\log \overline{G}(x)) + \delta \log \mu_G,$$

jolloin yhtälön (5.11) integraalille pätee

$$\int_0^x \frac{h(t)}{e_G(t)} dt = \underbrace{A(s) + \delta \log \mu_G}_{=O(1)} + o(\log \overline{G}(x)).$$

Sijoituksella $r(x) = \overline{F}(x)/\overline{G}(x)$ yhtälöstä (5.11) saadaan ratkaistua yhtäsuuruus

$$\begin{aligned}
-\log \overline{F}(x) &= -\log \overline{G}(x) - \overbrace{\log \left(\frac{\mu_F}{\mu_G} \right)}^{\in \mathbb{R}} - \log \left(\frac{e_G(x)}{e_F(x)} \right) + o(-\log \overline{G}(x)) + O(1) \\
&= -\log \overline{G}(x) - \log \left(\frac{e_G(x)}{e_F(x)} \right) + o(-\log \overline{G}(x)) + O(1).
\end{aligned} \tag{5.16}$$

Termin $o(-\log \overline{G}(x))$ avulla saadaan yhtälön (5.16) asymptoottinen käyttäytyminen määrättyä yhtälössä (5.15) olevia ala- ja ylärajoja hyödyntämällä. Tällöin arvioimalla ter-

mi $o(-\log \bar{G}(x))$ alaspäin nolaksi saadaan

$$-\log \bar{F}(x) \geq -\log \bar{G}(x) - \log\left(\frac{e_G(x)}{e_F(x)}\right) + O(1),$$

josta termillä $-\log \bar{G}(x) > 0$ puolittain jakamalla saadaan

$$\frac{-\log \bar{F}(x)}{-\log \bar{G}(x)} \geq 1 + \frac{\log\left(\frac{e_G(x)}{e_F(x)}\right)}{\log \bar{G}(x)} - \frac{O(1)}{\log \bar{G}(x)}.$$

Kun $x \rightarrow \infty$, niin tällöin $-\log \bar{G}(x) \rightarrow \infty$ ja oletuksen (i) mukaan $\log(e_G(x)/e_F(x)) \rightarrow 0$. Tällöin rajankäynnin $x \rightarrow \infty$ jälkeen saadaan epäyhtälö

$$\lim_{x \rightarrow \infty} \frac{-\log \bar{F}(x)}{-\log \bar{G}(x)} \geq 1. \quad (5.17)$$

Arvioimalla yhtälössä (5.16) olevaa termiä $o(-\log \bar{G}(x))$ ylöspäin saadaan

$$\begin{aligned} -\log \bar{F}(x) &\leq -\log \bar{G}(x) - \log\left(\frac{e_G(x)}{e_F(x)}\right) - c' \log \bar{G}(x) - O(1) \\ &= -\log\left(\frac{e_G(x)}{e_F(x)}\right) - \log \bar{G}(x)(1 + c') - O(1), \end{aligned}$$

josta jakamalla puolittain termillä $-\log \bar{G}(x)$ saadaan

$$\frac{-\log \bar{F}(x)}{-\log \bar{G}(x)} \leq \frac{\log\left(\frac{e_G(x)}{e_F(x)}\right)}{\log \bar{G}(x)} + 1 + c' + \frac{O(1)}{\log \bar{G}(x)}.$$

Raja-arvon $x \rightarrow \infty$ ottamisen jälkeen jäljelle jää epäyhtälö

$$\lim_{x \rightarrow \infty} \frac{-\log(\bar{F}(x))}{-\log(\bar{G}(x))} \leq 1 + c'.$$

Koska vakio c' saa olla mikä tahansa mielivaltaisen pieni positiivinen reaaliluku, niin tällöin rajalla $c' \rightarrow 0$ raja-arvolle saadaan yläraja

$$\lim_{x \rightarrow \infty} \frac{-\log(\bar{F}(x))}{-\log(\bar{G}(x))} \leq 1. \quad (5.18)$$

Yhdistämällä epäyhtälöiden (5.17) ja (5.18) antamat ala- ja ylärajat saadaan pääteltyä

$$\lim_{x \rightarrow \infty} \frac{-\log \bar{F}(x)}{-\log \bar{G}(x)} = 1 \quad \Leftrightarrow \quad -\log \bar{F}(x) \sim -\log \bar{G}(x).$$

Ekvivalenssin $-\log \bar{F}(x) \sim -\log \bar{G}(x)$ pohjalta tiedämme myös että kaikilla $\varepsilon > 0$ on olemassa $x_0 > 0$ jonka jälkeen kaikilla $x \geq x_0$ epäyhtälöt

$$1 - \varepsilon < \frac{-\log \bar{F}(x)}{-\log \bar{G}(x)} < 1 + \varepsilon$$

pätevät. Termillä $-\log \bar{G}(x)$ kertomisen ja logaritmin laskusääntöjen käyttämisen jälkeen

yllä oleva epäyhtälö saadaan ilmaistua muodossa

$$\overline{G}(x)^{1-\varepsilon} < \overline{F}(x) < \overline{G}(x)^{1+\varepsilon}.$$

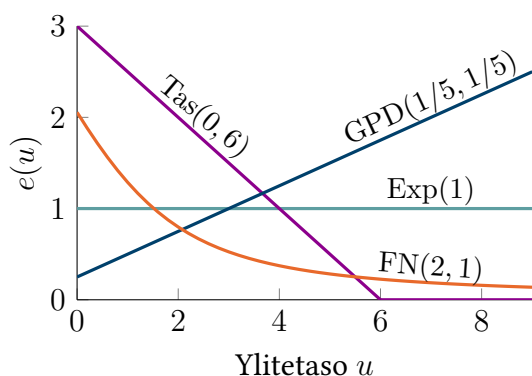
Täten yllä olevan epäyhtälön mukaan jokaisella $\varepsilon > 0$ tarpeeksi suurilla $x \geq x_0$ häntä-funktiolle $\overline{F}(x)$ saadaan ylä- ja alarajat häntäfunktion $\overline{G}(x)^{1\pm\varepsilon}$ arvoja hyödyntämällä, joka voidaan ilmaista myös muodossa $\overline{F}(x) = \overline{G}(x)^{1+o(1)}$. \square

Täten ehtojen (i) ja (ii) pätiessä ja odotusarvojen μ_F ja μ_G ollessa äärellisiä pätee asymp-toottinen yhtäsuuruus $-\log(\overline{F}(x)) \sim -\log(\overline{G}(x))$. Täten jos $\overline{F}(x)$ on tuntemattoman ja-kauman aineistosta arvioitu häntäfunktion ja $\overline{G}(x)$ tunnetun jakauman analyttinen häntä-funktio, niin tällöin proposition 5.9 mukaan suurilla x häntäfunktion $\overline{F}(x)$ voidaan korvata häntäfunktiolla $\overline{G}(x)$, sillä tästä aiheutuvat virheet ovat asymp-totiikan perusteella pieniä. Tästä on hyötyä varsinkin laskennallisissa tilanteissa, sillä aineistosta arvioitu häntäfunktion $\overline{F}(x)$ ei välttämättä anna yhtä tarkkaa tietoa jakauman hännästä kuin tunnettu analyttinen häntäfunktion $\overline{G}(x)$.

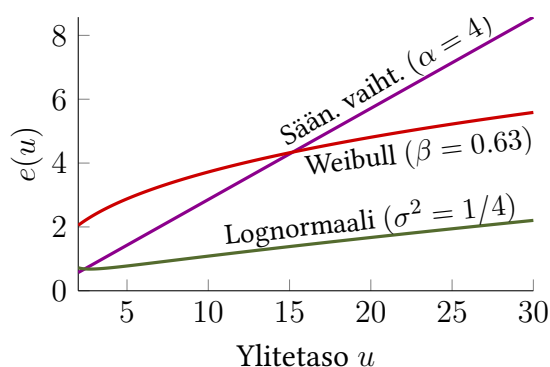
Proposition vaatiman (ii) ehdon pätevyyden varmistamiseen artikkelissa (Hall & Well-ner, 1981) esitettiin ehto $\liminf_{x \rightarrow \infty} e_G(x) > 0$. Ehdon pätiessä tulo $e_G(x)\overline{G}(x)$ voi olla nolla ja logaritmi $-\log(e_G(x)\overline{G}(x))$ ääretön pelkästään häntäfunktion $\overline{G}(x)$ toimesta. Täl-löin voidaan valita vakio $c > 0$ jolla yläraja $-c \log(\overline{G}(x))$ pätee termille $-\log(e_G(x)\overline{G}(x))$, sillä jälkimmäinen termi ei voi mennä äärettömäksi ilman että ensimmäinen termi menisi äärettömäksi.

5.3 | Hännän tyypin määrääminen

Jo johdannossa on kerrottu mahdollisuudesta tehdä jakaumapäätelmiä kuvan 1.1 kaltaisia kuvaajia hyödyntämällä. Esityslauseen 4.14 ja proposition 5.9 pohjalta tiedämme, että jos keskiylitysfunktion kuvaaja muistuttaa jonkin tunnetun jakauman keskiylitysfunktiota, niin tällöin myös taustalla olevat jakaumat ovat melkein samoja. Voidaanko tämän lisäksi tehdä vielä yleisempiä päätelmiä jakauman tyypistä jos keskiylitysfunktio ei muistuta mitään tunnettua keskiylitysfunktiota? Seuraavassa kuvassa on esitetty luvuissa 4 ja 5.1 laskettuja keskiylitysfunktioita vierekkäin silmämääräisen tarkastelun helpottamiseksi.



(a) Luvun 4 keskiylitysfunktiot



(b) Luvun 5.1 keskiylitysfunktiot

Kuva 5.2: Tutkielmassa laskettujen kevyt- ja paksuhäntäisten jakaumien keskiylitysfunktioiden kuvaajat.

Yksi kuvasta 5.2 huomattava asia on keskiylitysfunktioiden asymptoottinen käyttäytyminen. Kuvan 5.2a jakaumista tasa-, eksponentti- ja normaalijakauman keskiylitysfunktiot ovat kevythäntäisiä, kun taas yleistetty Pareto-jakauma on paksuhäntäinen jakauma. Vastaavasti kuvan 5.2b paksuhäntäisten jakaumien keskiylitysfunktiot ovat kasvavia.

Erot keskiylitysfunktioiden asymptoottisessa käyttäytymisessä näyttäisivätkin johtuvan jakauman hännän paksuudesta. Kaikkien rajatta kasvavien keskiylitysfunktioiden jakaumien häntä on paksumpi kuin eksponenttijakauman häntä, ja kaikkien nolaa kohti suppevien jakaumien häntä on kevyempi kuin eksponenttijakauman häntä. Eksponenttijakauma toimiikin kuvassa 5.2a rajatapauksena, joka erottelee kevyt- ja paksuhäntäiset jakaumat toisistaan. Myöhempää käyttöä varten on hyödyllistä yhdistää kuvasta 5.2 tehdyt empiiriset huomiot seuraavaksi matemaattisesti perustelemattomaksi yhteenvedoksi:

Tutkituilla paksuhäntäisillä jakaumilla keskiylitysfunktio kasvaa rajatta ja tutkituilla kevythäntäisillä jakaumilla keskiylitysfunktio on laskeva. Eksponentiaalijakauman keskiylitysfunktio on vakio.

Samoihin huomioihin keskiylitysfunktion käyttäytymisestä on päädytty myös kirjallisuudessa (Embrechts ym., 1997; McNeil ym., 2005; Ghosh & Resnick, 2010; Asmussen & Steffensen, 2020). Yhteenvetoa hyödyntämällä voidaan tehdä nopea aineistopohjainen tarkastelu joka kertoo tulisiko mallinnuksessa käyttää kevyt- tai paksuhäntäistä jakaumaa. Tämän lisäksi yleisessä käytössä olevien jakaumien keskiylitysfunktiot ovat kirjallisuudessa tunnettuja, ja niitä hyödyntämällä on mahdollista tehdä tarkempia päätelmiä aineiston taustalla olevasta jakaumasta. Keskiylitysfunktioita hyödyntävään aineistopohjaiseen päättelyyn tutustutaan tarkemmin luvussa 6.

6 | Sovelluksia aineistoihin

Edellisissä luvuissa esitettiin miten keskiylitysfunktio käyttäytyy ja millaisia yleisiä tuloksia sille pätee. Käytännön hyötyä varten näitä tietoja tulisi pystyä hyödyntämään myös aineistopohjaisessa päättelyssä, missä positiivisarvoisen havaintoaineiston X_1, \dots, X_n generoinut jakauma ei välttämättä ole selvillä. Tällöin ei voida suoraan laskea analyttistä esitysmuotoa keskiylitysfunktiolle, vaan on käytettävä empiiristä jakaumaa. Miten tällaisessa tilanteessa keskiylitysfunktiota voi hyödyntää aineistopohjaisessa päättelyssä?

Tämän luvun tavoitteena on esittää miten lukujen 4 ja 5 tuloksia voi hyödyntää aineistopohjaisessa mallinnuksessa, käyttäen esimerkkinä R:ssä generoituja aineistoja sekä luvuissa 6.2 ja 6.3 vapaasti saatavilla olevia aineistoja. Aineistopohjaisen työskentelyn vuoksi on määriteltävä tapa laskea keskiylitysfunktio havaintoaineiston X_1, \dots, X_n pohjalta. Tähän tarkoitukseen kirjallisuudessa käytetään *otoskeskiylitysfunktiota*.

Määritelmä 6.1. (Otoskeskiylitysfunktio) Havaintoaineistosta X_1, X_2, \dots, X_n saadaan laskettua *otoskeskiylitysfunktio* $e_n(u)$ käyttämällä kaavaa

$$e_n(u) = \frac{1}{\#\{j : X_j > u\}} \sum_{j: X_j > u} (X_j - u),$$

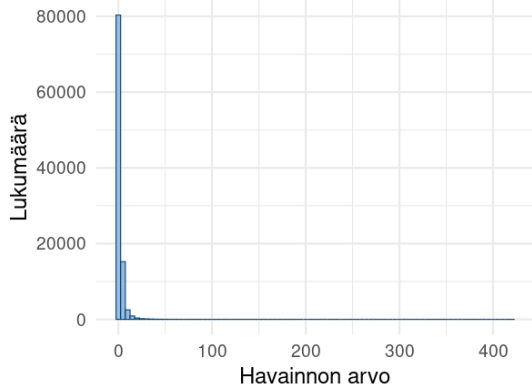
missä $\#\{j : X_j > u\}$ on tason u ylittävien havaintojen lukumäärä ja $\sum_{j: X_j > u} (X_j - u)$ on tason u ylittävien havaintojen X_j ja u välisten etäisyyksien summa.

Otoskeskiylitysfunktion implementointi eri laskentaohjelmistoille ei ole haastavaa. Esimerkiksi R:llä toteutuksen voi tehdä itse (Koodi B.2) tai hyödyntämällä jo olemassa olevia paketteja, kuten *ReIns*-paketissa (Reynkens & Verbelen, 2020) löytyvää komentoa *MeanExcess*. Kun otoskeskiylitysfunktion arvot on laskettu jokaisella havainnolla X_i , niin tämän jälkeen pistejoukon

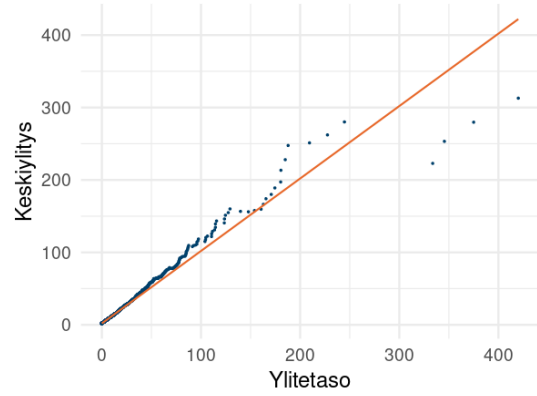
$$\{(X_i, e_n(X_i)) \mid i \in \{1, \dots, n\}\}$$

pohjalta piirretään otoskeskiylityskuvaaja. (Ebrechts ym., 1997)

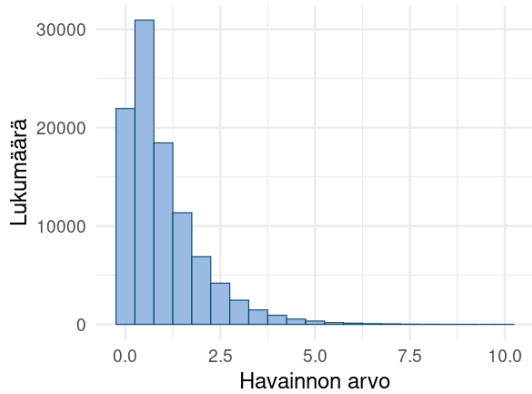
Otoskeskiylitysfunktion käyttäytymisen tarkasteluun on parasta lähestyä satunnaisesti generoidun aineiston avulla. Tällöin tiedämme yksikäsitteisesti jakauman F josta havainnot X_1, \dots, X_n on saatu. Käyttämällä jakauman häntäfunktiota $F(x)$ voidaan laskea myös analyttinen keskiylitysfunktio, ja piirtämällä otos- ja analyttinen keskiylitysfunktio samaan kuvaajaan voidaan vertailla mitä mahdollisia eroja näiden kahden käyttäytymisessä on. Kuvassa 6.1 yleistetystä Pareto-jakaumasta parametrein $\xi = 1/2$ ja $\beta = 1$, eksponenttijakaumasta parametrilla $\lambda = 1$ ja taitetusta normaalijakaumasta $FN(2, 1)$ on simuloitu 100 000 havaintoa, ja näiden pohjalta on piirretty simuloitujen jakaumien otos- ja analyttiset keskiylitysfunktiot. Otoskeskiylitysfunktioiden piirtämisessä on sovellettu koodia B.4.



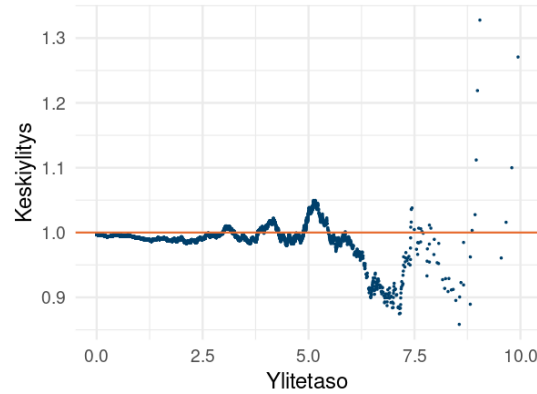
(a) Yleistetystä Pareto -jakaumasta $GPD(1/2, 1)$ generoidun havaintoaineiston histogrammi, otoskoko $n = 100000$.



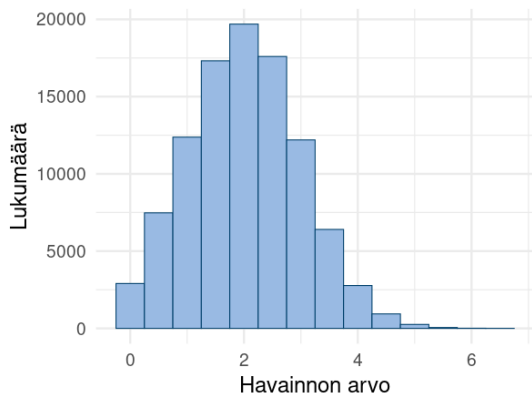
(b) Yleistetyn Pareto -jakauman analyttinen (esimerkki 4.6) ja generoidun aineiston pohjalta laskettu otoskeskiylitysfunktio.



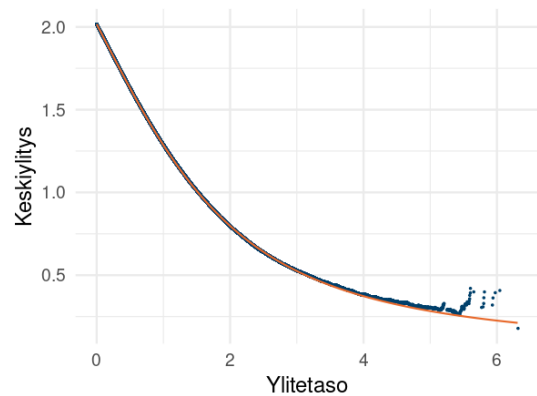
(c) Eksponenttijakaumasta $Exp(1)$ generoidun havaintoaineiston histogrammi, otoskoko $n = 100000$.



(d) Eksponenttijakauman $Exp(1)$ analyttinen (esimerkki 4.5) ja generoidun aineiston pohjalta laskettu otoskeskiylitysfunktio.



(e) Taitetusta normaalijakaumasta $FN(2, 1)$ generoidun havaintoaineiston histogrammi, otoskoko $n = 100000$.



(f) Taitetun normaalijakauman $FN(2, 1)$ numeerisesti ja generoidun aineiston pohjalta laskettu otoskeskiylitysfunktio.

Kuva 6.1: Yleistetystä Pareto-, eksponentti- ja taitetusta normaalijakaumasta generoitujen havaintoaineistojen ($n = 100000$) frekvenssihistogrammit ja näiden pohjalta laskettujen analyttisten (oranssi) ja otoskeskiylitysfunktioiden (sininen) kuvaajat.

Tarkastelemalla kuvan 6.1 histogrammeja ja otoskeskiylitysfunktioita vierekkäin on mahdollista nähdä miten otoskeskiylitysfunktio reagoi havaintojen määrän muutukseen. Jokaisessa esimerkin tilanteessa havaintojen arvojen kasvaessa havaintojen lukumäärä vähenee, mikä muodostaa omia haasteitaan aineiston analysoinnissa suurilla arvoilla. Tällaisissa tilanteissa on järkevintä käyttää luvussa 3.3 esitettyä kaikkia tason $u > 0$ ylittäviä havaintoja hyödyntävää tason ylittävät huiput -menetelmää, johon myös keskiylitysfunktiot pohjautuvat ylitejakaumien kautta.

Aineiston pohjalta laskettu otoskeskiylitysfunktio ei kuitenkaan muodosta täysin teoreettista vastinetta vastaavia kuvaajia. Esimerkiksi eksponenttijakauman tilanteessa (Kuva 6.1d) arvon $u = 3$ jälkeen otoskeskifunktio ei ole enää teoreettisen vastineensa kaltainen vakio, vaan sen kuvaaja alkaa olemaan ykkösen molemmilla puolilla. Samaa käyttäytymistä on havaittavissa myös yleistetyn Pareto-jakauman otoskeskiylitysfunktion kuvaajassa (Kuva 6.1b), jossa ylitetaso $u = 125$ jakaa otoskeskiylitysfunktion käyttäytymisen kahteen eri tilanteeseen; arvoilla $u \leq 125$ kasvu on lineaarista mutta arvoilla $u > 125$ kasvun lineaarisuus ei ole enää yhtä selkeästi havaittavissa. On kuitenkin huomautettava kuinka lähellä otoskeskiylitysfunktio on teoreettista vastinettaan havaintojen määrän ollessa suuri, kuten taitetun normaalijakauman tilanteessa (Kuva 6.1f) on selkeiten nähtävissä. Vasta havaintojen lukumäärän vähentyessä radikaalisti otoskeskiylitysfunktion käyttäytyminen alkaa selvästi eroamaan teoreettisesta vastineestaan.

Otoskeskiylitysfunktioiden kuvaajat ovat muutenkin harvoin täysin samankaltaisia kuin niiden teoreettiset vastineet. Tämä näkyy selkeimmin suurilla ylitetasoilla, sillä tällöin tason $u > 0$ ylittävistä havainnoista lasketaan keskiarvo jonka mahdollisesta arvosta ei ole mitään takeita. Tämä johtuu siitä, että keskiarvon laskemisessa käytetään vähäistä määrää pisteitä, ja yksittäisellä poikkeavalla havainnolla voi olla suuri vaikutus keskiarvoon. Tämän vuoksi keskiylityskuvaajia piirtäessä on perusteltua jättää muutama suurin havainto piirtoaineiston ulkopuolelle. (McNeil ym., 2005) Kuvan 6.1 keskiylitysfunktioiden piirtämisessä on jätetty kolme suurinta havaintoa huomioimatta, ja tästä eteenpäin kaikissa tämän luvun otoskeskiylityskuvaajissa noudatetaan samaa tapaa.

Kuvan 6.1 esimerkkien lisäksi on myös hyödyllistä tietää miten kirjallisuudessa keskiylitysfunktioita hyödynnetään eri aineistojen pohjalla jakaumaa mallinnettaessa. Seuraavassa luvussa luodaankin yleiskatsaus keskiylitysfunktion käyttöön eri mallinnustilanteissa.

6.1 | Otoskeskiylitysfunktion hyödyntäminen

Ennen keskiylitysfunktioiden piirtämistä havaintoaineiston pohjalta on palautettava mieleen luvussa 3 esitetyt teoreettiset vaatimukset. Kyseisessä luvussa esitettiin matemaattinen perusta ääriarvoteorialle, mutta käytännön sovelluskohteissa ei aina ole mahdollista hyödyntää analyttisiä esitysmuotoja ja vesitiiviitä oletuksia jakaumasta F . Sovelluskohteissa käsitellään havaintoaineistoja, joiden pohjalta lasketaan tunnuslukuja ja piirretään eri otoksiin pohjautuvia kuvaajia. Tällöin onkin hyvä kysyä millä tavalla äärimmäiset havainnot näkyvät aineistoissa (Embrechts ym., 1997). Ääriarvoteoria antaa meille kuitenkin hyviä työkaluja harvinaisten tapahtumien mallintamiseen, ja jos niitä haluaa käyttää on pyrittävä edes jollain tavalla varmistamaan luvussa 3 esitettyjen oletusten voimassaolo.

Luvun 3 oletuksista oleellisin oletus on havaintojen X_1, \dots, X_n samoin jakautuminen ja riippumattomuus sekä oletus $F \in MDA(H_\xi)$ jollain $\xi \in \mathbb{R}$, sillä tässä työssä esitetty ääriarvoteoria pohjautuu juurikin näiden oletuksien hyödyntämiseen. Näiden oletusten ollessa voimassa on myös perusteltua käyttää mallinnuksessa määritelmässä 3.11 esitetty-

jä Gumbelin, Fréchetin ja Weibullin jakaumia, sillä melkein kaikki yleisessä käytössä olevat jakaumat kuuluvat näiden kolmen vaikutuspiiriin (Embrechts ym., 1997). Havaintojen X_1, \dots, X_n samoin jakautuneisuus ei myöskään ole triviaalia, vaikka havainnot olisivatkin havaittu samasta prosessista eri vuosien aikana. Jos havainto X_i esimerkiksi kuvastaa maksettua vakuutuskorvausta, niin eri vuosien maksut voi inflaatiokorjata samalle tasolle, jolloin on perustellumpaa olettaa havaintojen X_1, \dots, X_n olevan samoin jakautuneita.

Oletusten ollessa kunnossa voidaan vihdoin alkaa käyttämään keskiylitysfunktioita aineiston analysoinnissa. Keskiylitysfunktioiden käytetyin ominaisuus sovelluskohteissa on luvussa 5.3 esitetyn yhteenvedon hyödyntäminen jakauman hännän paksuuden määrittämisessä. Tällöin kasvavan keskiylitysfunktion tilanteessa mallinnuksessa kannattaa valita käytettäväksi paksuhäntäinen jakauma, ja vastaavasti keskiylitysfunktion laskiessa kevyt-häntäinen jakauma. Kun mallinnuksessa käytettävän jakauman hännän paksuus on valittu, niin tällöin voidaan keskittyä tarkemmin itse jakauman tyyppiin ja parametrien arviointiin. (Embrechts ym., 1997)

Tämän lisäksi keskiylitysfunktioita voi hyödyntää tarkemmin jakauman hännän määrittämisessä. Yksi tapa liittyy suoraan Pickands–Balkema–de Haanin lauseen (Lause 3.17) hyödyntämiseen. Kyseisen lauseen oletusten täyttyessä tarpeeksi suurilla x ylitejakaumaa $F_u(x)$ voi approksimoida yleistetyn Pareto-jakauman kertymäfunktiota $G_{\xi, \beta(u)}(x)$ käyttäen. Esimerkissä 4.6 laskettu yleistetyn Pareto-jakauman keskiylitysfunktio

$$e(u) = \frac{\beta + \xi u}{1 - \xi} = \frac{\xi}{1 - \xi} u + \frac{\beta}{1 - \xi} \quad (6.2)$$

on lineaarinen, ja myös kuvassa 6.1b esiintyvä yleistetyn Pareto-jakauman otoskeskiylitysfunktio vaikuttaisi olevan verrattaen lineaarinen. Täten jos keskiylitysfunktio vaikuttaisi olevan lineaarinen, niin tällöin on perusteltua yrittää sovittaa yleistettyä Pareto-jakaumaa ylitejakaumaksi. (Hall & Wellner, 1981; Embrechts ym., 1997; McNeil ym., 2005)

Otoskeskiylitysfunktiot eivät kuitenkaan ole aina täysin lineaarisia. Usein jakauman alussa keskiylitysfunktio voi käyttäytyä eri tavalla kuin jakauman loppupäässä, jolloin voi olla järkevää pyrkiä mallintamaan jakaumaa F käyttämällä kahta täysin eri jakaumaa F_1 ja F_2 . Tällaista mallia sanotaan *sekamalliksi*. (Hogg & Klugman, 1984)

Määritelmä 6.3. (Sekajakauma) Olkoot $F_1(x)$ ja $F_2(x)$ kahden eri jakauman kertymäfunktioita ja ρ todennäköisyys että havaittu tapahtuma tuli jakaumasta F_1 . Tällöin *sekajakauman* kertymäfunktio $F(x)$ saadaan kaavalla

$$F(x) = \rho F_1(x) + (1 - \rho) F_2(x). \quad (6.4)$$

Usein äärimmäisiä ilmiöitä mallinnettaessa selkeä enemmistö havainnoista keskittyy pieniin arvoihin ja äärimmäisiä ilmiöitä kuvastavat suuret havainnot ovat vähäisiä. Tämän vuoksi jakauma F_1 voidaan valita kuvastamaan jakaumaa F välillä $[0, u)$ ja jakauma F_2 kuvastamaan ylitejakaumaa F_u välillä $[u, \infty)$, jolloin jakauman arviointi suoritetaan kahdessa eri osassa. Yksi tapa arvioida parametria ρ aineiston pohjalta on laskea suhteellisesti kuinka monta havaintoa kaikista havainnoista on välillä $[0, u)$, jolloin saamme laskennassa käytettävän arvion $\hat{\rho}$. Kertymäfunktion $F_1(x)$ arviointi voidaan usein tehdä havaintojen suuren lukumäärän vuoksi melko tarkasti *empiiristä kertymäfunktiota* käyttäen. (Hogg & Klugman, 1984)

Määritelmä 6.5. (Empiirinen kertymäfunktio) Olkoot X_1, \dots, X_n riippumattomia ja samoin jakautuneita havaintoja jakaumasta F . Tällöin jakauman F empiirinen kertymäfunktio pisteessä u on

$$F_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq u\}}.$$

Yksi mahdollinen tapa laskea empiirinen keskiylitysfunktio R:ssä on hyödyntää peruskomentoa *ecdf* tai implementoida se itse (Koodi B.3). Jos jollain välillä on tarpeeksi havaintoja jakaumasta F , niin empiirinen kertymäfunktio voi antaa käytettävän arvion kyseisen välin jakaumasta. On kuitenkin muistettava että jatkuvaa jakaumaa mallinnettaessa diskreetti $F_n(x)$ ei tule ikinä täysin vastaamaan jatkuvaa vastinettaan $F(x)$, jonka vuoksi olisi hyvä pyrkiä muodostamaan jatkuva arvio empiirisen kertymäfunktion pohjalta (Hogg & Klugman, 1984).

Yhtälön (6.4) mukaista sekajakaumaa sovellettaessa on kuitenkin tehtävä perusteltavissa oleva päätös mistä pisteestä u lähtien jakauma F hajotetaan jakaumiksi F_1 ja F_2 . Yksi tapa tehdä tämä on hyödyntää Pickands–Balkema–de Haanin lausetta (Lause 3.17) sekä esimerkiksi 4.6 laskettua yleistetyn Pareto-jakauman keskiylitysfunktiota. Tällöin tarkoituksena on pyrkiä valitsemaan sellainen ylitetaso $u \geq 0$ jonka jälkeen kaikilla $x \geq u$ otoskeskiylitysfunktio $e_n(x)$ on likimain lineaarinen, jolloin ylitejakaumaksi F_2 voidaan lauseen 3.17 mukaisesti valita yleistetty Pareto-jakauma sopivilla parametreilla. Tämänlaisella empiiristä kertymäfunktiota sekä yleistettyä Pareto-jakaumaa yhdistävällä lähestymistavalla voidaan saavuttaa hyvä arvio jakaumasta F , kunhan ylitetaso $u \geq 0$ on valittu tarpeeksi suureksi. (Embrechts ym., 1997)

Keskiylitysfunktioiden kuvaajien muotoja voi myös hyödyntää päättelyssä (Hogg & Klugman, 1984). Tässä tutkielmassa on laskettu joidenkin keskiylitysfunktioiden esitysmuotoja, ja lisää on laskettu kirjallisuudessa, kuten esimerkiksi taulukko 3.4.7 kirjasta (Embrechts ym., 1997). Esityslauseen (Lause 4.14) ja proposition 5.9 mukaan jos aineiston pohjalta piirretty keskiylitysfunktio muistuttaa jonkin tunnetun jakauman keskiylitysfunktiota, niin tällöin on perusteltua käyttää tunnettua jakaumaa ja sen tunnuslukuja arvioinnissa. Eri jakaumien keskiylitysfunktioiden muotojen tunnistamisen on sanottu olevan hyödyllinen taito äärimmäisiä arvoja mallintaessa. (Hall & Wellner, 1981)

Vaikka havaintoaineiston otoskoko n olisi erittäin suuri, niin tästä huolimatta joukossa $\{X_1, \dots, X_n\}$ on aina olemassa suurin mahdollinen havainto $m = \max(X_1, \dots, X_n)$. Tästä johtuen tiedämme vain sen, että jakauman F kantaja on ainakin väli $[0, m]$, mutta emme tiedä kuinka paljon pidemmälle kantaja jatkuu pisteen m jälkeen. Nykyään varsinkin vakuutus- ja rahoituskonteksteissa jakauman F oletetaan olevan luvussa 5.1 esiteltyjen aliekspontiaalisten jakaumien tyyppinen, jonka seurauksena välin $[0, m]$ havaintoja hyödyntämällä on myös pääteltävä miten F käyttäytyy välillä (m, ∞) . Tämänlainen välillä (m, ∞) jakauman F käyttäytymisen päättely ei enää perustu havaintoaineistoon vaan se on havaintoaineiston ekstrapolaatiota, jonka pätemisen perustelu ei ole enää triviaalia (Asmussen & Steffensen, 2020). Ääriarvoteoria muodostaa kuitenkin tukevan perustan jota hyödyntämällä voidaan pyrkiä muodostamaan arvioita myös erittäin suurien havaintojen todennäköisyyksistä. Suuret havaintoaineistot $\{X_1, \dots, X_n\}$ sisältävät myös äärimmäisiä havaintoja joita voidaan hyödyntää jakauman mallinnuksessa. (Embrechts ym., 1997)

On myös muistettava että keskiylitysfunktio ei ole ainut menetelmä mitä voi hyödyntää aineiston tuottaneen jakauman päättelyssä. Kuvassa 6.1 esiintyvistä histogrammeista voidaan nähdä mihin ja miten havaintojen arvot keskittyvät eri kohdille välille $[0, \infty)$. Toinen kirjallisuudessa usein käytetty työkalu on *kvanttiilikuvio*, jossa empiirisen havaintoaineiston

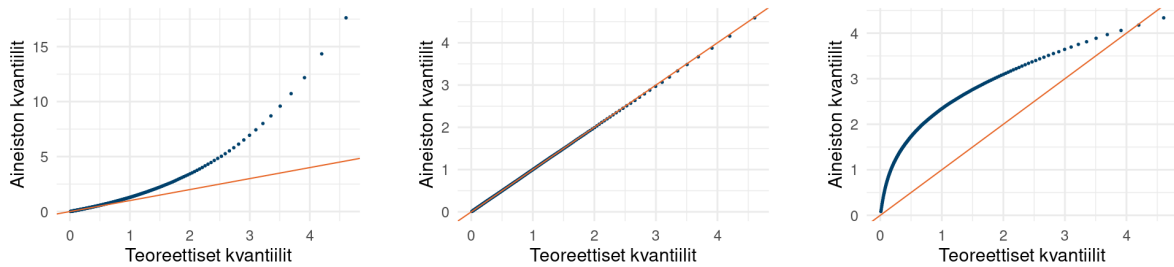
ja teoreettisen jakauman kvantiilit piirretään samaan kuvaajaan. Teoreettiset ja empiiriset kvantiilit $q \in [0, 1]$ saadaan yhtälöistä

$$Q(q) = \inf\{x \in \mathbb{R} : q \leq F(x)\} \quad \text{ja} \quad Q_n(q) = \inf\{x \in \mathbb{R} : q \leq F_n(x)\},$$

jonka jälkeen kvantiilikuvion pisteet saadaan piirtämällä joukko

$$\{Q(q), Q_n(q)\}$$

argumentin q suhteen yli välin $[0, 1]$. Kvantiilikuvion pisteiden käyttäytyminen havainnollistaa näiden kahden jakauman kvantiilien välistä yhteyttä. Jos pistejoukon approksimaatiokäyrän kasvu poikkeaa suoran $x = y$ kaltaisesta lineaarisesta kasvusta, niin tällöin teorettinen jakauma ei vastaa empiiristä jakaumaa (McNeil ym., 2005). Kuvassa 6.2 on piirretty kuvan 6.1 generoiduille aineistoille kvantiilikuviot teoreettisen jakauman $\text{Exp}(1)$ suhteen sekä suora $x = y$ helpottamaan visuaalista tarkastelua.



- (a) Yleistystä Pareto - jakaumasta $\text{GPD}(1/2, 1)$ generoitujen havaintojen kvantiilit teoreettisia $\text{Exp}(1)$ kvantiileita vastaan.
- (b) Eksponenttijakaumasta $\text{Exp}(1)$ generoitujen havaintojen kvantiilit teoreettisia $\text{Exp}(1)$ kvantiileita vastaan.
- (c) Taitetusta normaalijakaumasta $\text{FN}(2, 1)$ generoitujen havaintojen kvantiilit teoreettisia $\text{Exp}(1)$ kvantiileita vastaan.

Kuva 6.2: Kuvassa 6.1 generoitujen havaintoaineistojen kvantiilien vertaaminen jakauman $\text{Exp}(1)$ teoreettisia kvantiileita vastaan. Siniset pisteet vastaavat kvantiileita ja oranssi viiva suoraa $x = y$.

Kuvan 6.2 kvantiilikuvioista voidaan nähdä miten kvantiilikuvioiden pisteet käyttäytyvät verrattaessa empiirisiä kvantiileita tunnetun jakauman kvantiileita vastaan. Empiiristen ja teoreettisten kvantiileiden ollessa samat kvantiilikuvio on kuvan 6.2b kaltainen, jossa kvantiilikuvion pisteet asettuvat suoralle $x = y$. Yleistetyn Pareto-jakauman tilanteessa (Kuva 6.2a) empiiristen kvantiilien arvot nousevat nopeammin kuin teoreettisten kvantiilien arvot, joka kvantiilikuviossa näkyy pistejoukon kiihtyvänä kasvuna. Tällöin voimme päätellä, että empiirisen aineiston taustalla on paksuhäntäisempi jakauma kuin verrokkina käytetty eksponenttijakauma. Kuvasta 6.2c on havaittavissa empiirisen aineiston kvantiilien nopea kasvu välillä $(0, 1)$, jonka jälkeen kasvunopeus hidastuu. Tämä viittaisi siihen, että empiirillä aineistolla on pisteen $x = 1$ ympäristössä todennäköisyysmassakeskittymä, ja pisteen $x = 4$ jälkeen todennäköisyysmassa on vähemmän kuin eksponenttijakaumalla.

Hyödyntämällä monia eri työkaluja aineiston analysoinnissa, kuten histogrammeja ja kvantiilikuvioita, saadaan empiirisen aineiston pohjalta muodostettua parempia arvioita todellisesta jakaumasta. Histogrammeja hyödyntämällä voidaan nähdä onko aineistossa havaittavissa arvojen keskittymistä, keskiylitysfunktiolla voidaan päätellä jakauman kevyt- tai paksuhäntäisyys ja kvantiilikuvioilla voidaan tarkastella vastaako empiirisen jakauman

kvantiilit jonkin tunnetun jakauman kvanttileita. Kirjallisuudessa keskiylitysfunktion sano-taankin olevan hyvä lisätyökalu aineiston visuaaliseen tarkasteluun, sillä keskiylitysfunktiot pystyvät kuvastamaan jakauman hännän tyyppiä paremmin kuin esimerkiksi histogrammit (Hall & Wellner, 1981; Hogg & Klugman, 1984). Tutkielmassa mainittujen tapojen lisäksi muita mahdollisia aineistopohjaisen päättelyn työkaluja ja tilastollisia testejä on käsitelty kirjallisuudessa (Hogg & Klugman, 1984; Coles, 2001; Kleiber & Kotz, 2003).

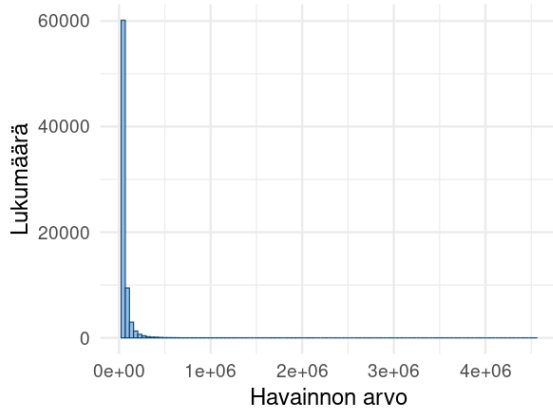
Ääriarvoteoriaan pohjautuvia menetelmiä hyödyntämällä, kuten keskiylitysfunktioita ja yleistettyä Pareto-jakaumaa käyttämällä, saadaan sovitettua malli joka kuvastaa häntäta-pahtumien vakavuuksia paremmin kuin pelkästään empiiristä kertymäfunktiota hyödyntä-mällä saataisiin. Mallin sovittamisen jälkeen sovitettua jakaumaa voidaan empiirisen jakau-man kaltaisesti hyödyntää sovelluskohteissa tehtävissä laskennassa. Esimerkiksi finanssia-lalla sovitetulla jakaumalla voidaan helpommin laskea VaR-lukuja, simuloida erilaisia tap-piotilanteita ja laskea ekvivalenssiperiaatteen mukaisia nettokertamaksuja. (Embrechts ym., 1997) Tunnetun jakauman avulla edellä mainittuja laskuja voidaan helpommin laskea hyö-dyntämällä monissa laskentaohjelmistoissa olevia komentoja, jotka usein vaativat pelkäs-tään arvioituja parametreja laskentaa varten. Tällöin myös uuden aineiston pohjalta päivi-tetyt parametrit voidaan myöhemmin syöttää jo valmiina oleviin komentoihin, sillä yksittäi-sen parametrin muuttaminen tilastollisessa mallissa on nopeampaa kuin koko havaintoai-neiston uudelleen käsittely. (Hogg & Klugman, 1984)

6.2 | SOA Group Medical Insurance Large Claims

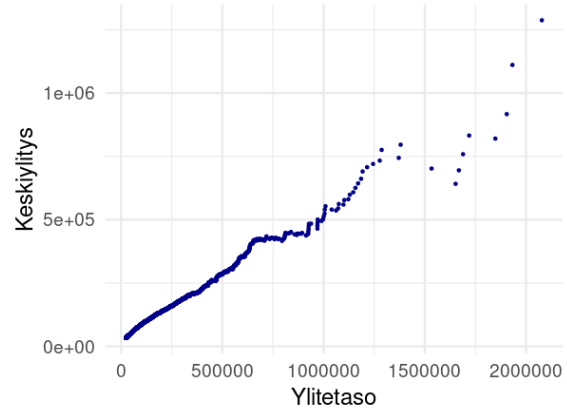
SOA Group Medical Insurance Large Claims on yhdysvaltalaisen aktuaariyhdistyksen *Society of Actuaries (SoA)* julkaisema aineisto yhdysvaltalaisten työnantajien tarjoamien sairaskulu-vakuutuksien 25 000 \$ suuruuden ylittävien korvauksien arvoista vuodelta 1991. Aineistossa havaintoja on yhteensä 75 789 kappaletta ja sen lähteenä on R-paketti *ReIns* (Reynkens & Verbelen, 2020).

Koska aineisto sisältää havaintoja pelkästään 25 000 \$ ja sitä suuremmista korvausmää-ristä, niin tällöin voidaan olettaa että välin $[0, 25\,000]$ jakaumalle on jo jollain tavalla muo-dostettu jakauma-arvio. Tällöin kyseessä on luvussa 3.1 esitetyn ylitejakauman arviointi ta-solla $u = 25\,000$ \$, johon tässä tutkielmassa esitetty ääriarvoteoria soveltuu mainiosti. Voim-me myös helpotuksen vuoksi olettaa kaikkien havaintojen olevan riippumattomia ja samoin jakautuneita, jotta luvussa 3 esitettyä ääriarvoteoriaa voidaan hyödyntää ongelmitta. Tämä oletus ei välttämättä todellisuudessa päde, sillä erilaisilla aloilla on erilaiset riskit sairastua ja vammautua niin työ- kuin vapaa-ajalla.

Aineiston tarkastelun voi aloittaa piirtämällä kuvan 6.3 kaltaisesti histogrammin ja otos-keskiylitysfunktion kuvaajan. Histogrammista voi pyrkiä katsomaan miten aineiston arvot jakautuvat reaalityökaluakselille ja otoskeskiylitysfunktion kuvaajasta voi pyrkiä päättämään jakauman hännän paksuutta. Pelkästään jo näitä kahta työkaluja käyttämällä voidaan mah-dollisesti löytää sopiva aineistoon sovitettava malli.



(a) SoA-aineiston frekvenssihistogrammi ($n = 75\,789$).



(b) SoA-aineiston otoskeskiylitysfunktion kuvaaja.

Kuva 6.3: SoA-aineiston histogrammi ja otoskeskiylitysfunktion kuvaaja.

Kuvan 6.3a histogrammista on havaittavissa samanlaista arvojen keskittymistä pienille arvoille, mutta havaintoja on myös jonkin verran erittäin suurilla arvoilla. Tämä käyttäytyminen viittaa paksuhäntäiseen jakaumaan, ja piirretty histogrammi muistuttaakin kuvan 6.1a kaltaista paksuhäntäisen yleistetyn Pareto-jakauman histogrammia. Täten histogrammin pohjalta aineistoon olisi perusteltua sovittaa paksuhäntäistä jakaumaa.

Keskiylitysfunktion kuvaaja (Kuva 6.3b) vaikuttaa olevan silmämääräisesti melkein täysin lineaarinen tasoon 1 250 000 \$ asti, jonka jälkeen havaintojen lukumäärän vähentyessä otoskeskiylitysfunktion arvot alkavat lisääntyvissä määrin poikkeamaan lineaariselta suoralta. Tämä käyttäytyminen on odotettavissa suurilla ylitetasoilla, sillä tällöin käytännössä otetaan keskiarvo vähäisestä määrästä havaintoja. Otoskeskiylitysfunktion arvojen kasvu on kuitenkin pääosin lineaarista, jonka mukaan SoA-aineistoon voitaisiin sovittaa yleistetty Pareto-jakaumaa yhtälön (6.2) perusteella.

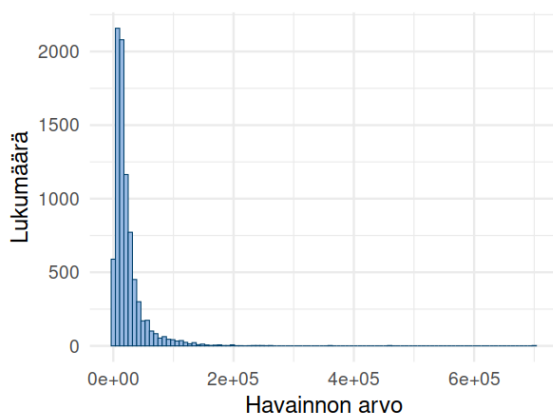
Koska kuvan 6.3a histogrammi viittaa paksuhäntäiseen jakaumaan ja kuvan 6.3b otoskeskiylitysfunktion kuvaaja lineaarisuuden perusteella näyttäisi vastaavan yleistetyn Pareto-jakauman keskiylitysfunktiota, niin tällöin on perusteltua kokeilla sovittaa aineistoon yleistettyä Pareto-jakaumaa. Tätä päätöstä tukee myöskin lause 3.17, jonka mukaan tarpeeksi suurilla u ylitejakauma suppenee kohti yleistettyä Pareto-jakaumaa. Jakauman sovitus voidaan tehdä esimerkiksi R:ssä paketin *ReIns* komentoa *GPDfit* käyttämällä, joka antaa jakauman määrittelyssä tarvittavien parametrien estimaattorit $\hat{\xi}$ ja $\hat{\beta}(u)$. Parametrien selvittämisen jälkeen voidaan sovitetun jakauman sopivuutta testata ja hienosäätää esimerkiksi kvantiilikuvia käyttämällä.

Jakauman parametrien hienosäädön jälkeen voidaan siirtyä jakauman aiheuttaman riskin tarkasteluun. Luvun 4.3 riskimittoja hyödyntämällä voidaan pyrkiä muodostamaan parempaa käsitystä mahdollisista maksettavista sairaskulukorvauksista. VaR-luvun ja odotetun tappion avulla saadaan laskettua jakauman riskiä eri kvantiilitasoilla $\alpha \in (0, 1)$ vastaavia tunnuslukuja, joita voidaan myöhemmin hyödyntää yrityksen sisäisessä riskienhallinnassa ja viranomaisraportoinnissa. Sovitetun jakauman avulla voidaan myös helpommin simuloida erilaisia tilanteita, ja tarkastella niiden pohjalta miten yrityksen taloudellinen tilanne reagoi korvausvaatimusten lukumäärän muutokseen.

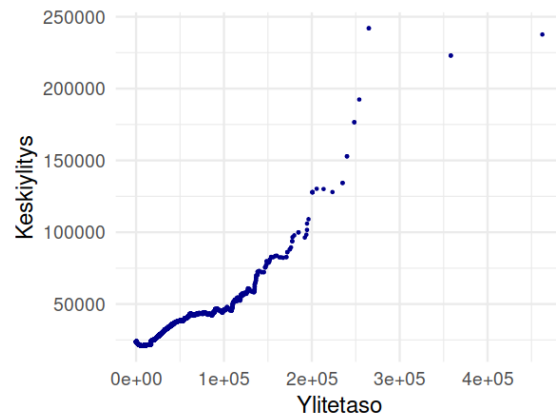
6.3 | Norwegian automobile insurance losses

Norwegian automobile insurance losses (norauto) sisältää asiakastietoja norjalaisen vakuutusyhtiön myöntämien liikennevakuutusten tiedoista yhden vuoden ajalta. Aineisto sisältää tiedot yhteensä 183 999 sopimuksesta. Jokaisen sopimuksen kohdalla on tarjolla tarkempia tietoja sopimuksesta, kuten esimerkiksi vakuutuksenottajan asuinalue, sukupuoli, liikenneympäristön vilkkaus sekä tiedot vuoden aikana maksetuista korvauksista. Näiden asiakkaiden joukossa korvauksia on maksettu vuoden aikana 8 444 vakuutetulle. Korvausmäärät on ilmoitettu norjan kruunuissa. Aineiston lähteenä tässä tutkielmassa on R-paketti (Dutang & Charpentier, 2020).

Aineisto mahdollistaisi asiakkaiden jakamisen eri asiakkuusluokkiin esimerkiksi asuinpaikan ja ajettujen kilometrien perusteella, jolloin voitaisiin jokaiselle luokalle mallintaa oma tappiojakauma ja tätä hyödyntämällä määritellä jokaiselta kuluttajalta perittävä luokakohtainen vakuutusmaksu. Tässä esimerkissä havaintojen on kuitenkin oletettu olevan riippumattomia ja samoin jakautuneita tarkastelun yksinkertaistamisen vuoksi. Keskiylitysfunktion käytön vaadittavien oletusten käsittelyn jälkeen voidaan aineiston tulkinta aloittaa piirtämällä histogrammi ja otoskeskiylitysfunktion kuvaaja.



(a) Norauto-aineiston frekvenssihistogrammi ($n = 8\,444$).



(b) Norauto-aineiston otoskeskiylitysfunktion kuvaaja.

Kuva 6.4: Norauto-aineiston histogrammi ja otoskeskiylitysfunktion kuvaaja.

Aineiston histogrammissa (Kuva 6.4a) voidaan nähdä miten valtaosa havainnoista keskittyy välille $[0, 200\,000 \text{ NOK})$, jonka jälkeen suuria havaintoja on harvakseltaan. Tämä havaintojen keskittyminen välin $[0, 700\,000 \text{ NOK})$ ensimmäiselle kolmannekselle viittaisi jakauman häntään olevan kuvan 6.1 histogrammien pohjalta kevyempi kuin yleistetyn Pareto-jakauman häntä, mutta samalla paksumpi kuin eksponenttijakauman häntä. Täten histogrammin perusteella kyseessä olisi mahdollisesti jokin paksuhäntäinen jakauma, mutta täyttä varmuutta sovitettavasta jakaumasta ei pystytä sanomaan pelkästään tämän pohjalta.

Otoskeskiylitysfunktion kuvaajaa (Kuva 6.4b) hyödyntämällä voidaan pyrkiä valitsemaan tarkemmin alustavasti sovitettavat jakaumat. Välillä $[0, 100\,000 \text{ NOK})$ otoskeskiylitysfunktio vaikuttaisi kasvavan logaritmisesti, joka muistuttaa kuvassa 5.2b esiintyvän Weibullin jakauman keskiylitysfunktion käyttäytymistä. Välillä $[100\,000, 200\,000 \text{ NOK})$ kasvu on verrattain lineaarista, joka viittaa yhtälön (6.2) mukaisesti yleistettyyn Pareto-jakaumaan. 200 000

kruunun ylittävät arvot käyttäytyvät erittäin epäsäännöllisesti. Kasvu on alussa melkein pystysuoraa, jonka jälkeen kolme viimeistä arvoa ovatkin samantasoisia. Havaintojen vähyden vuoksi 200 000 kruunun ylittävän jakauman valinnaista ei voida muodostaa otoskeskiylitysfunktion perusteella tarkkaa kuvaa, vaikka kasvava trendi viittaakin paksuhäntäisyyteen.

Otoskeskiylitysfunktion vaihtelevan käyttäytymisen perusteella aineistoon voitaisiin sovittaa määritelmän 6.3 mukaista sekajakaumaa, missä jakaumaa pyrittäisiin mallintamaan käyttämällä kahta tai kolmea eri jakaumaa. Välille $[0, 100\ 000\ \text{NOK})$ voitaisiin pyrkiä sovittaa Weibullin jakaumaa edellämainitun logaritmisin kasvun perusteella. Jäljelle jäävää tason 100 000 NOK ylittävää jakaumaa voidaan mallintaa kahdella eri tavalla; joko sovitamalla väleille $[100\ 000, 200\ 000\ \text{NOK})$ ja $[200\ 000\ \text{NOK}, \infty)$ erilliset jakaumat tai välille $[100\ 000\ \text{NOK}, \infty)$ yksi yhtenäinen jakauma. Lineaarisen kasvun perusteella hyvä ehdokas näille kahdelle sovitukselle olisi yleistetty Pareto-jakauma. Jakaumien alustavan sovittamisen jälkeen voidaan sovituksen sopivuutta tarkastella esimerkiksi kuvan 6.2 kaltaisia kvanttilikuvioita hyödyntämällä.

7 | Lähteet

- Asmussen, S. & Steffensen, M. (2020). *Risk and Insurance*. Probability Theory and Stochastic Modelling. Springer.
- Billingsley, P. (1995). *Probability and measure*. Wiley series in probability and mathematical statistics. Wiley, New York, NY, 3. edition.
- Bingham, N. H., Teugels, J. L., & Goldie, C. M. (1987). *Regular Variation*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer series in statistics. Springer, Bristol.
- Dutang, C. & Charpentier, A. (2020). *CASdatasets: Insurance datasets*. R package version 1.0-11.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events: for Insurance and Finance*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg.
- Ghosh, S. & Resnick, S. (2010). A discussion on mean excess plots. *Stochastic Processes and their Applications*, 120(8):1492 – 1517.
- Grimmett, G. R., Stirzaker, D., ym. (2001). *Probability and random processes*. Texts from Oxford University Press. Oxford University Press, 3rd edition.
- Hall, W. & Wellner, J. A. (1981). Mean residual life. *Statistics and related topics*, pages 169–184.
- Hogg, R. V. & Klugman, S. A. (1984). *Loss distributions*. Wiley series in probability and mathematical statistics. Wiley, New York.
- Kleiber, C. & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*, volume 470. John Wiley & Sons.
- Leppisaari, M. (2013). *Katastrofi- ja finanssiriskin mittaamisesta ääriarvoteoriaa soveltaen*. Aalto-yliopisto.
- McNeil, A. J., Embrechts, P., & Frey, R. (2005). *Quantitative risk management : concepts, techniques and tools*. Princeton series in finance. Princeton University Press, Princeton.
- Peters, G. & Shevchenko, P. (2015). *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*. Wiley Handbooks in Financial Engineering and Econometrics. Wiley.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Reynkens, T. & Verbelen, R. (2020). *ReIns: Functions from “Reinsurance: Actuarial and Statistical Aspects”*. R package version 1.0.10.

Woods, F. S. (1926). *Advanced calculus: a course arranged with special reference to the needs of students of applied mathematics*. Ginn.

Yamai, Y. & Yoshihara, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, 29(4):997–1015.

A | Liitteet

Määritelmä A.1. (Samantyyppiset jakaumat) Satunnaismuuttujien X ja Y jakaumien sanotaan olevan *samantyyppisiä* jos ja vain jos on olemassa sellaiset vakiot $\alpha > 0$, $\beta \in \mathbb{R}$, että yhtälö

$$X \stackrel{d}{=} \alpha Y + \beta$$

pätee.

Lause A.2. (Tyyppittäinen suppeneminen) Olkoot X, Y, X_1, X_2, \dots satunnaismuuttujia ja $b_n > 0, \beta_n > 0$ ja $a_n, \alpha_n \in \mathbb{R}$ vakioita. Oletetaan että suppeneminen

$$\frac{X_n - a_n}{b_n} \xrightarrow{d} X$$

pätee. Tällöin suppeneminen

$$\frac{X_n - \alpha_n}{\beta_n} \xrightarrow{d} Y \tag{A.3}$$

pätee jos ja vain jos

$$\lim_{n \rightarrow \infty} \frac{b_n}{\beta_n} = b \in [0, \infty) \quad \text{ja} \quad \lim_{n \rightarrow \infty} \frac{a_n - \alpha_n}{\beta_n} = a \in \mathbb{R}.$$

Jos (A.3) pätee, niin tällöin

$$Y \stackrel{d}{=} bX + a$$

ja a, b ovat ainoat vakiot jotka toteuttavat yllä olevan yhtälön.

Todistus. Lauseen todistus löytyy kirjasta (Billingsley, 1995, luku 2.14). □

Määritelmä A.4. (Pareto-jakauma) Parametrein $\alpha > 0$ ja $k > 0$ Pareto-jakautuneen satunnaismuuttujan X jakaumamerkintä, kertymäfunktio ja odotusarvo ovat

$$X \sim \text{Par}(\alpha, k), \quad F(x) = 1 - \left(\frac{k}{x}\right)^\alpha \quad \text{ja} \quad \mathbb{E}(X) = \begin{cases} \infty, & \alpha \leq 1, \\ \frac{k\alpha}{\alpha-1}, & \alpha > 1. \end{cases}$$

Pareto-jakauman kantaja on väli $x \in [k, \infty)$.

Määritelmä A.5. (Taitettu normaalijakauma) Olkoon $X \sim N(\mu, \sigma)$. Tällöin satunnaismuuttujaa $Y = |X|$, jolloin merkitään $Y \sim \text{FN}(\mu, \sigma)$, kutsutaan taitetuksi normaalijakaumaksi, missä parametria μ sanotaan sijaintiparametriksi ja σ muoto parametriksi. Taitetun normaalijakauman kantaja on väli $[0, \infty)$ ja sen tiheysfunktio on muotoa

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

Määritelmä A.6. (Weibullin jakauma) Parametrein $\alpha > 0$ ja $\beta > 0$ Weibull-jakautuneen satunnaismuuttujan X jakaumamerkintä, kertymäfunktio ja odotusarvo ovat

$$X \sim \text{Wei}(\alpha, \beta), \quad F(x) = 1 - \exp\{-\alpha x^{-\beta}\} \quad \text{ja} \quad \mathbb{E}(X) = \alpha^{-1} \Gamma(1 + \beta^{-1}),$$

missä $\Gamma(t)$ on gammafunktio. Weibullin jakauman kantaja on väli $[0, \infty)$.

Määritelmä A.7. (Lognormaalijakauma) Satunnaismuuttuja X on log-normaalijakautunut parametrein $\mu \in \mathbb{R}$ ja $\sigma^2 > 0$ jos ja vain jos satunnaismuuttuja $Y = \ln(X)$ on normaalijakautunut. Log-normaalijakauman kantaja on väli $(0, \infty)$, ja jakaumamerkintä ja odotusarvo ovat

$$X \sim \text{LogN}(\mu, \sigma^2) \quad \text{ja} \quad \mathbb{E}(X) = \exp\{\mu + \sigma^2/2\}.$$

Määritelmä A.8. (Ordo-notaatio) Olkoot $f(x)$ ja $g(x)$ kaksi funktiota reaalityyppisillä. Tällöin kyseisten funktioiden asymptoottista järjestystä sekä suppenemisnopeuksia kun $x \rightarrow \infty$ voidaan merkitä seuraavilla tavoilla:

- Notaatiolla $O(g(x))$, jolloin tarkoitetaan joukkoa

$$O(g(x)) = \{f(x) \mid \exists c > 0 \exists x_0 > 0 \forall x \geq x_0 : 0 \leq f(x) \leq cg(x)\},$$

missä funktio $g(x)$ on asymptoottisesti funktion $f(x)$ yläraja.

- Notaatiolla $o(g(x))$, jolloin tarkoitetaan joukkoa

$$o(g(x)) = \{f(x) \mid \forall c > 0 \exists x_0 > 0 \forall x \geq x_0 : 0 \leq f(x) \leq cg(x)\},$$

missä funktio $g(x)$ on asymptoottisesti funktion $f(x)$ aito yläraja.

Lisätietoa Ordo-notaatiosta, sen ominaisuuksista ja käytöstä kirjallisuudessa löytyy lähteestä (Peters & Shevchenko, 2015, luku 3.1).

Lause A.9. (Karamatan lause) Olkoon $L \in \mathcal{R}_0$ lokaalisti rajoitettu välillä $[x_0, \infty)$ jollain $x_0 \geq 0$. Tällöin

- a) kun $\alpha > -1$:

$$\int_{x_0}^x t^\alpha L(t) dt \sim (\alpha + 1)^{-1} x^{\alpha+1} L(x), \quad x \rightarrow \infty,$$

- b) kun $\alpha < -1$:

$$\int_x^\infty t^\alpha L(t) dt \sim -(\alpha + 1)^{-1} x^{\alpha+1} L(x), \quad x \rightarrow \infty.$$

Todistus. Lauseen todistus löytyy kirjasta (Bingham ym., 1987, luku 1.5.7). □

Lemma A.10. (Leibnizin integroimissääntö) Olkoon $f(x, t)$ ja sen osittaisderivaatta $\frac{\partial}{\partial x} f(x, t)$ jatkuvia funktiota muuttujan t suhteen, ja jatkuvia muuttujan x suhteen tasolla (x, t) , mukaan lukien alueet $a(x) \leq t \leq b(x)$ ja $x_0 \leq x \leq x_1$. Oletetaan myös, että funktiot $a(x)$ ja $b(x)$ ovat molemmat jatkuvasti derivoituvia välillä $x_0 \leq x \leq x_1$. Tällöin välillä $x_0 \leq x \leq x_1$ pätee

$$\frac{d}{dx} \left(\int_{a(x)}^{b(x)} f(x, t) dt \right) = f(x, b(x)) \cdot \frac{d}{dx} b(x) - f(x, a(x)) \cdot \frac{d}{dx} a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x, t) dt.$$

Todistus. Lemman todistus löytyy kirjasta (Woods, 1926, luku VI:60). □

B | Tutkielmassa käytetyt R-koodit

Koodi B.1. Kuvissa 4.1 ja 5.2a esiintyvän taitetun normaalijakauman keskiylitysfunktion pisteiden laskemisessa käytetty koodi.

```
library(VGAM)

Q <- function(x) 1-pfoldnorm(x, mean = 2, sd = 1)
Q_int <- function(u) integrate(Q,u,Inf)$value

Norm_MEF_num <- function(u) Q_int(u)/Q(u)
Norm_MEF_num_vec <- function(u) sapply(u, Norm_MEF_num)

x <- seq(0, 7, by = 0.2)
y <- Norm_MEF_num_vec(x)
```

Koodi B.2. Yksi tapa laskea otoskeskiylitysfunktion $e_n(u)$ arvo pisteessä u :

```
mef <- function(u, data) {
  sum(subset(data, data > u)-u)/sum(data > u)
}
```

Koodi B.3. Yksi tapa laskea empiirisen kertymäfunktion $F_n(x)$ arvo pisteessä x :

```
emp_cdf <- function(x, data) {
  sum(data <= x)/length(data)
}
```

Koodi B.4. Yksi tapa piirtää havaintoaineiston pohjalta otoskeskiylityskuviot R:ää käyttämällä. Tällä koodilla on piirretty kuva 6.3b.

```
library(ReIns)
library(ggplot2)

mef <- function(u, data) {
  sum(subset(data, data > u)-u)/sum(data > u)
}

data(soa)
soa <- sort(soa$size)

mef_soa <- function(x) mef(x, soa)
soa_ex <- unlist(lapply(soa, mef_soa), use.names=FALSE)

data_soa <- data.frame(x = head(soa,-3), y = head(soa_ex,-3))

ggplot(data_soa, aes(x,y)) +
  labs(x = "Ylitetaso", y="Keskiylitys") +
  geom_point(color="darkblue", fill="lightblue") +
  theme_minimal()
```