

Antti Lempinen

**YLEISTETYN PARETO-JAKAUMAN KÄYTTÖ  
TANSKASSA SATTUNEIDEN TULIPALOJEN  
AIHEUTTAMIEN VAKUUTUSKORVAUSTEN  
MALLINTAMISESSA**

Diplomityö

Tekniikan ja luonnontieteiden tiedekunta

Tarkastajat: Yliopistonlehtori Simo Ali-Löytty ja Tenure track -professori Pasi Raumonen

Syyskuu 2023

## TIIVISTELMÄ

Antti Lempinen: Yleistetyn Pareto-jakauman käyttö Tanskassa sattuneiden tulipalojen aiheuttamien vakuutuskorvausten mallintamisessa

Diplomityö

Tampereen yliopisto

Teknis-luonnontieteellinen DI-ohjelma

Syyskuu 2023

---

Ympäriamme esiintyy paljon ilmiöitä, joita ei voida kuvata normaalijakauman avulla, sillä sen häntäpäissä esiintyneet havainnot voivat vaikuttaa merkittävästi saatuun lopputulokseen. Tässä työssä esitelläänkin tähän ongelmaan kehitelty paksuhäntäinen jakauma ja kerrotaan muutamia sen käyttökohteita reaali maailmassa. Paksuhäntäinen jakauma määritellään matemaattisesti ja perehdytään vielä tarkemmin yhteen sen ryhmään kuuluvista jakaumista, yleistettyyn Pareto-jakaumaan. Yleistetty Pareto-jakaumaa käytetään lopuksi mallintamaan aineistoa, joka sisältää dataa Tanskassa vuodesta 1980 vuoteen 1990 toteutuneista tulipaloista aiheutuneista palovakuutuskorvauksista. Mallintaminen sisältää jakauman parametrien estimoinnin hyödyntämällä R-ohjelmistoa. Mallinnuksessa huomataan, että yleistetty Pareto-jakauma sopii kuvaamaan kyseistä dataa, joten mallintamisen ja estimoinnin jälkeen datalle lasketaan vielä VaR (Value at Risk) ja ES (Expected shortfall) luvut 5%, 2,5% ja 1% todennäköisyyksille.

Avainsanat: Pareto, GPD, paksuhäntäinen jakauma

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

## ABSTRACT

Antti Lempinen: Using a generalised Pareto distribution to model insurance claims from fires in Denmark

Master of Science Thesis

Tampere University

Master's Programme in Science and Engineering

September 2023

---

There are many phenomena around us that cannot be described by a normal distribution, as observations at the tail-end of the distribution can have a significant impact on the final result. In this paper, we present a heavy-tailed distribution developed to address this problem and describe some of its applications in the real world. A heavy-tailed distribution is defined mathematically and one of the distributions in its class, the generalised Pareto distribution, is explored in more detail. Finally, the generalised Pareto distribution is used to model the Danish fire insurance claims data (contains data from 1980 to 1990). The modelling includes estimation of the parameters of the distribution using R-software. The modelling process shows that the generalised Pareto distribution is suitable to describe this data, so after modelling and estimation, the VaR (Value at Risk) and ES (Expected shortfall) are calculated for 5%, 2,5% and 1% probabilities.

Keywords: Pareto, GPD, heavy-tailed distribution

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

## ALKUSANAT

Vuonna 2017 alkanut yliopistourani opiskelijana on tullut päätökseen. Haluaisin kiittää opiskelujeni sujumisesta lähipiiriäni, mukaan lukien opiskelukavereitani. Lisäksi haluaisin lähettää erityiskiitokset tyttöystävälleni tuesta ja kannustamisesta sekä kandi- että diplomityöni ohjanneelle Simo Ali-Löytylle, joka ohjasi töideni tekoa oikeaan suuntaan ja mahdollisti näiden loppuun saattamisen.

Olen edelleen erittäin onnellinen opiskelualani valinnasta ja pääsen nyt vihdoinkin myös hyödyntämään sen minulle suomiani oppeja ja astumaan niistä ovista joita se on minulle avannut.

Tampereella, 1. syyskuuta 2023

Antti Lempinen

## SISÄLLYSLUETTELO

1.	Johdanto . . . . .	1
2.	Paksuhäntäisten jakaumien käyttö mallinnuksessa . . . . .	3
2.1	Paksuhäntäisten jakaumien käyttö ilmatoriskin mallintamisessa . . . . .	3
2.2	Paksuhäntäisten jakaumien käyttö muualla reaali maailmassa . . . . .	4
3.	Paksuhäntäinen jakauma . . . . .	6
3.1	Paksuhäntäisen jakauman määritelmä . . . . .	6
3.2	Yleistetty Pareto-jakauma . . . . .	10
3.3	Muita paksuhäntäisiä jakaumia . . . . .	12
3.3.1	Cauhchy-jakauma. . . . .	13
3.3.2	Burr-jakauma . . . . .	14
4.	Yleistetyn Pareto-jakauman käyttö mallintamisessa . . . . .	17
4.1	Tanskalainen tulipalodata . . . . .	17
4.2	Datan mallinnus yleistetyllä Pareto-jakaumalla . . . . .	18
4.2.1	Mallin hyvyyden ja parametrien estimointi . . . . .	19
4.2.2	Mallin sovitus dataan . . . . .	20
4.2.3	VaR- ja ES-lukujen laskeminen . . . . .	23
5.	Yhteenveto . . . . .	27
	Lähteet . . . . .	29
	Liite A: Koodi . . . . .	32

## LYHENTEET JA MERKINNÄT

$:=$	Määritelmä (luetaan "on määritelmän mukaan")
$\arctan$	Arkustangentti
DKK	Tanskan kruunu (rahayksikkö)
ES	Expected shortfall
$f(x)$	Tiheysfunktio
$F(x)$	Kertymäfunktio
GPD	Generalized Pareto distribution eli yleistetty Pareto-jakauma
$\bar{F}$	Häntäfunktio
$\hat{F}$	"Hattu" yläpuolella = estimaatti
$\inf$	Infimum eli suurin alaraja
$\int$	Integraali
$\approx$	Likimain yhtäsuuri
$M(s)$	Momentit generoiva funktio
MLE	Maximum Likelihood Estimation eli suurimman uskottavuuden estimaatti
$\mathbb{R}$	Reaaliluvut
$R(x)$	Hazard funktio
$\sin$	Sini
$\sinh$	Hyperbolinen sini
$\Sigma$	Summa
$\sup$	Supremum eli pienin yläraja
$\tan$	Tangentti
$\tanh$	Hyperbolinen tangentti
TAU	Tampereen yliopisto (engl. Tampere University)
$\amalg$	Tulo
TUNI	Tampereen korkeakouluyhteisö (engl. Tampere Universities)
URL	Verkkosivun osoite (engl. Uniform Resource Locator)

VaR

Value at Risk

# 1. JOHDANTO

Maailmassa on paljon tilanteita, joissa on syytä pohtia skenaarioita, jotka poikkeavat huomattavastikin historian valossa toteutuneista tilanteista, sillä niiden sattuessa vaikutus on merkittävä. Esimerkiksi vakuutusyhtiöt eivät voi täysin sivuuttaa tilannetta, joissa vakuutuksen kohteena oleva kiinteistö tuhoituisi täysin myrskyn takia, vaikka historian valossa kyseisellä alueella poikkeuksellisen kovien myrskyjen esiintyminen olisikin todella harvinaista. Tällaisissa tilanteissa mallintamisen apuna käytetään usein paksuhäntäisiä jakaumia. Paksuhäntäiset jakaumat antavat näille harvinaisille, mutta merkittäville tapauksille normaalijakaumaa korkeammat todennäköisyydet, joten ne tulevat mallinnuksessa paremmin huomioitua.

Paksuhäntäisiä jakaumia on useita erilaisia ja niitä käytetään useassa eri tieteenalassa. Tässä työssä on tarkoitus esitellä muutamia niiden käyttökohteita ja määritellä niille tyypillisiä ominaispiirteitä. Lisäksi työssä esitellään yksi yleisesti käytössä olevista paksuhäntäisistä jakaumista, yleistetty Pareto-jakauma vielä tarkemmin. Yleistetty Pareto-jakauma on valittu tarkempaan tarkasteluun sillä työssä keskitytään mallintamaan maastopaloista aiheutuneita palovakuutuskorvauksia ja sen käytöstä on esimerkkejä [8], [27] ilmastoon liittyvissä mallinuksissa. Maastopalojen aiheuttamat palovakuutuskorvaukset on valittu mallinnuksen kohteeksi, sillä maastopalot ja ilmastoasiat ovat tällä hetkellä ajankohtaisia.

Työssä yleistetty Pareto-jakauma tuodaan käytäntöön sovittamalla sitä dataan, joka sisältää Tanskassa vuodesta 1980 vuoteen 1990 toteutuneita tulipaloista aiheutuneita palovakuutuskorvauksia. Työn tutkimuskysymykset voidaan muotoilla seuraavasti:

- Onko Tanskassa sattuneiden tulipalojen aiheuttamia palovakuutuskorvauksia sisältävä data paksuhäntäisesti jakautunut?
- Voidaanko kyseistä dataa mallintaa yleistetyn Pareto-jakauman avulla?

Tämän lisäksi työn lopussa lasketaan konkreettisia, rahamääräisiä, palovakuutuskorvausten suuruuksia valituilla todennäköisyyksillä.



Työn ensimmäiseen tutkimuskysymykseen pyritään saamaan vastaus piirtämällä datasta R-ohjelmiston avulla keskimääräisen ylityksen funktio ("mean excess function"). Yleistetyn Pareto-jakauman sopivuuden testaamiseksi datasta estimoidaan suurimman uskottavuuden estimaatin (maximum likelihood estimation, MLE) avulla jakaumalle sopivimmat parametrit ja tarkastellaan tällä tavoin saadun jakauman sopivuutta QQ-kuvaajan avulla. Kyseiset menetelmät ovat estimoinnissa ja testaamisessa yleisesti käytettyjä.

## **2. PAKSUHÄNTÄISTEN JAKAUMIEN KÄYTTÖ MALLINNUKSESSA**

Paksuhäntäiset jakaumat sopivat kuvaamaan tilanteita, joissa jotkin tapaukset ovat harvinaisia, mutta niiden vaikutus on poikkeuksellisen suuri. Esimerkkejä näistä ovat poikkeuksellisen voimakkaiden luonnonilmiöiden aiheuttamat vahingot (esimerkiksi maanjäristys, hurrikaani, tulivuorenpurkaus, tulva tai maastopalo), eloonjäämis- tai toipumisajat lääketeieteessä tai useamman toisistaan riippumattoman joukon yhteismallintaminen. Tapaukset, jotka ovat yleisiä ja lähellä odotusarvoa eivät puolestaan sovi hyvin paksuhäntäiseen mallintamiseen.

### **2.1 Paksuhäntäisten jakaumien käyttö ilmatoriskin mallintamisessa**

Oletetaan, että pisin ihminen, jonka olet tähän päivän mennessä tavannut on tasan 2 metriä pitkä. Jonain päivänä tulet kuitenkin todennäköisesti kohtaamaan vieläkin pidemmän ihmisen. Kuinka todennäköisenä pidät, että tämä henkilö tulee olemaan kaksi kertaa yhtä pitkä kuin edellinen pisin kohtaamasi ihminen eli 4 metriä pitkä? Todennäköisyys on äärettömän pieni, oikeastaan nolla. Ajatellaanpa toinen tilanne: ennen vuotta 2005 suurin hurrikaanin aiheuttama tappio yhdisvalloissa oli 41,5 miljardia dollaria, vuonna 2011 hurrikaani Katrinan aiheuttamat tappiot olivat kuitenkin 91 miljardia dollaria. On siis tapahtunut tilanne, jossa historiaan viittavan datan pohjalta on sattunut yli kaksinkertainen kasvu odotettuun maksimiin. Ilmatoon liittyvät ääriarvotapahtumat ovatkin yksi esimerkki paksuhäntäisistä tapahtumista, kun taas ihmisen pituudet voidaan arvioida jakautuvat ohuthäntäisesti. [7]

Yllä olevassa esimerkissä huomataan, että varautuminen tulevaan, poikkeukselliseen tapahtumaan, voi olla yrityksille kriittistä. Mikäli vakuutusyhtiöt olivat arvioineet hurrikaanin aiheuttamia korvauksia paksuhäntäisesti, ne olivat varmasti jo hinnoitteluvaiheessa ottaneet riskin huomioon ja näin ollen siirtäneet taloudellista riskiä vakuutusentajille esimerkiksi korkeampien vakuutusmaksujen muodossa. Jos taas riskiä oli arvioitu ohuthäntäisesti, toteutuneet korvaukset ovat olleet yritykselle erittäin merkittävä kuluerä ja

pahimmassa tapauksessa ajaneet koko yrityksen konkurssiin.

Ilmastoriskit ovat erittäin suuria riskejä esimerkiksi vakuutusyhtiöille, joten niitä on pyritty mallintamaan jo melko pitkään. Käytettyjä malleja on useita ja niitä muokataan yhä yksityiskohtaisimmiksi ja kulloiseenkin tapahtumaan sopiviksi. Paksuhäntäisiä jakaumia käytetään ilmastomallinnuksessa yhä tänäkin päivänä, ja vaikka mallit ovatkin kehittyneet vuosien varrella, yksi yleinen tapa testaamiseen on kuitenkin melko vanha malli, yleistetty Pareto-jakauma [6].

Tässä työssä tullaan tarkastelemaan ilmastoriskin, tarkemmin maastopalojen, mallintamista. Maastopalot ovat maailmalla suuria vahinkoja aiheuttava luonnonilmiö, joten niitä on pyritty mallintamaan jo aikaisemminkin. Mallintamisessa ei ole havaittavissa yhtä oikeaa tapaa vaan eri tutkimuksissa on kuvattu vahinkoja eri mallien avulla. Esimerkiksi Yhdysvalloissa ennen vuosituhannen vaihtumista sattuneita vahinkoja on mallinnettu onnistuneesti niin Weibull-jakauman ([26] ja [21]) kuin lognormaalijakauman [29] avulla.

Reed ja McKelvey [26] tarkastelivat vuonna 2002 ilmestyneessä teoksessaan metsäpaaloissa palaneiden alueiden jakautumista. Data sisälsi havaintoja 1900-luvulta ja se oli kerätty kuudelta eri Pohjois-Amerikan alueelta. Työssä havaittiin, että mikäli palon leviämiseen vaikuttaneille tekijöille (esimerkiksi palaneen alueen ilmasto ja maasto) saatiin arvioitua oikeat parametrit, kolmen parametrin Weibull-jakauman avulla niiden jakautumista saatiin mallinnettua melko hyvin. Työssä palaneiden alueiden jakautumista testattiin kolmella eri mallilla, joista kolmen parametrin Weibull-jakauma tuotti tarkimman mallinnuksen etenkin suuria paloja sisältävien alueiden kohdalla.

Strauss, Bednar ja Mees [29] julkaisivat jo vuonna 1989 teoksen, jossa he pyrkivät mallintamaan Yhdysvaltojen länsiosassa sattuneiden maastopalojen kokoa. Työn tarkoituksena oli mallintaa harvoin, mutta pinta-alallisesti merkittävien, sattuneiden palojen vaikutusta kaikkien palojen kokonaisvaikutukseen. Työssä havaittiin, että palojen kokojen erittäin suuren vaihtelun vuoksi Pareto-jakauma antoi liian paksuhäntäisiä tuloksia ja parempi mallinnus tulos saatiin suoritettua lognormaalijakauman avulla.

## 2.2 Paksuhäntäisten jakaumien käyttö muualla reaali maailmassa

Ilmastoriskin mallintamisen lisäksi toinen merkittävä paksuhäntäisiä jakaumia hyödyntävä ala on lääketiede. On olemassa monia sairauksia, jotka vaativat vakavampaa hoitoa kuten esimerkiksi leikkauksia. Näiden hoitojen tarpeen arvioinnissa on otettava huomioon mahdolliset riskit esimerkiksi hoidon jälkeisessä selviytymisessä ja toipumisessa. Tämä-

kin riski on paksuhäntäinen, sillä vaikka suurin osa historian tapauksista olisikin sujunut hyvin, mahdollinen komplikaatio hoidon jälkeen on erittäin merkittävä ja sitä ei tule sivuuttaa vaikka sitä ei aikaisemmin olisikaan juuri tapahtunut. Tässäkin tapauksessa malleja on useita ja ne pitää sovittaa aina kulloisenkin tilanteen mukaan, mutta yksi yleisesti käytetty malli on Burr-jakauma, tarkemmin Burrin tyyppi XII. Tutkimuksessa on esimerkiksi havaittu, että satunnaisotoksella valitun 128 virtsarakon syöpäpotilaan remissioajat (toipumisaika) noudattaa Burr-jakaumaa [13].

Tilanteissa, joissa halutaan tutkia kahden eri, nolla odotusarvoisen normaalijakautuneen joukon, yhteistä jakaumaa käytetään usein Cauchy-jakaumaa. Jakaumaa käytetään monissa sovelluksissa kuten mekaanisessa sähköteoriassa, fysikaalisessa antropologiassa ja mittausongelmissa. Fysiikassa sitä kutsutaan usein myös Lorenzin jakaumaksi. Mallintamisen edellytyksenä on, että kaksi tutkittavaa datajoukkoa eivät ole toisistaan riippuvaisia. Sitä onkin käytetty onnistuneesti esimerkiksi mallintamaan tuberkuloosibakteerilla infektoidujen marsujen elossaoloajan ja Niilin vuotoisen veden virtaaman datojen yhteismallinnusta. [1]

### 3. PAKSUHÄNTÄINEN JAKAUMA

Kuten jo edellisessä kappaleessakin mainittiin paksuhäntäinen jakauma ottaa normaalijakaumaa paremmin huomioon jakaumien häntäarvot. Yleisenä kuvauksena voidaan myös pitää sitä, että mikäli jakauman häntä on paksumpi kuin eksponenttijakaumalla sen voidaan sanoa olevan paksuhäntäinen [4]. Tarkastellaan tässä kappaleessa nyt tarkemmin paksuhäntäisen jakauman määritelmää ja tutustutaan muutaman sen erikoistapauksen ominaisuuksiin hieman tarkemmin.

#### 3.1 Paksuhäntäisen jakauman määritelmä

Tässä työssä ollaan kiinnostuneita jakaumien (oikean) häntän ominaisuuksista eli jakauman ominaisuuksista, jotka riippuvat minkä tahansa muuttujan  $x$  osalta vain jakauman arvoista välillä  $(x, \infty)$ . Jatkossa on helpompi käsitellä funktioiden häntien ominaisuuksia omina funktioinaan. Satunnaismuuttujan  $X$  kertymäfunktioille  $F$  voidaan määrittää häntä-funktio  $\bar{F}$  seuraavasti

$$\bar{F}(x) = \mathbb{P}(X > x) = 1 - F(x), x \in \mathbb{R} \quad (3.1)$$

**Määritelmä 3.1.** Kertymäfunktioilla  $F$  on oikealle rajoittamaton kantaja, jos  $\bar{F}(x) > 0$  kaikilla muuttujan  $x$  arvoilla.

**Määritelmä 3.2.** Tiheysfunktio  $f$  on (oikealle) paksuhäntäinen, jos

$$\int_{\mathbb{R}} e^{\lambda x} f(x) dx = \infty \quad \text{kaikille } \lambda > 0 \quad (3.2)$$

**Määritelmä 3.3.** Tiheysfunktio  $f$  on ohuhäntäinen, jos

$$\int_{\mathbb{R}} e^{\lambda x} f(x) dx < \infty \quad \text{jollakin } \lambda > 0, \quad (3.3)$$

toisin sanoen jos se ei ole paksuhäntäinen.

Lauseessa 3.1 todistamme miten paksuhäntäisiä jakaumia voidaan määritellä ekvivalentisti usealla eri tavalla. Tätä todistusta varten teemme vielä seuraavat määritelmät:

**Määritelmä 3.4.** Kertymäfunktio  $F$  on paksuhäntäinen jos

$$\limsup_{x \rightarrow \infty} \bar{F}(x)e^{\lambda x} = \infty \quad \text{kaikille } \lambda > 0. \quad (3.4)$$

Toisin sanoen satunnaismuuttujan  $X$  kasvaessa kohti ääretöntä myös sen kertymäfunktion  $F$  häntäfunktion  $\bar{F}$  supistumisen pienin yläraja lähenee ääretöntä. Lause 3.1 osoittaa lisäksi, että jakauma on paksuhäntäinen, jos ja vain jos sen häntäfunktio on paksuhäntäisen funktio.

**Määritelmä 3.5.** Kertymäfunktion  $F$  funktiota  $R(x) := -\ln \bar{F}(x)$  kutsutaan jakauman *hazard funktioksi*. Jos hazard funktio on differentoituva, sen derivaattaa  $r(x) = R'(x)$  kutsutaan *hazard rate*:ksi. [12]

**Määritelmä 3.6.** Jatkuvan satunnaismuuttujan  $X$ , tiheysfunktioilla  $f(x)$ ,  $k$ :nnes momentti (mikäli se on äärellisenä olemassa) on muotoa

$$\mathbb{E}(X^k) = \int_{\mathbb{R}} x^k f(x) dx.$$

**Määritelmä 3.7.** Momentit generoiva funktio  $M : \mathbb{R} \rightarrow \mathbb{R}$  satunnaismuuttujalle  $X$  määritellään säännöllä

$$M(s) = \mathbb{E}(e^{sX}),$$

kaikilla  $s \in \mathbb{R}$ .

**Lause 3.1.** Tarkastellaan satunnaismuuttujaa  $X$ . Seuraavat väittämät ovat ekvivalentteja:

- (i)  $X$  on paksuhäntäinen.
- (ii) Momentit generoiva funktio  $M(s) := \mathbb{E}[e^{sX}] = \infty$  kaikille  $s > 0$ .
- (iii) Vastaava hazard funktio  $R$  toteuttaa  $\liminf_{x \rightarrow \infty} \frac{R(x)}{x} = 0$ .

*Todistus.* (i)  $\Rightarrow$  (ii). Oletetaan, että  $X$  on jatkuva ja paksuhäntäinen noudattaen kertymäfunktioita  $F$ . Määritelmän 3.4 mukaan tästä seuraa, että millä tahansa  $s > 0$  on olemassa aidosti kasvava sarja  $(x_k)_{k \geq 1}$  siten, että  $\lim_{k \rightarrow \infty} x_k = \infty$  ja täten

$$\lim_{k \rightarrow \infty} e^{sx_k} \bar{F}(x_k) = \infty. \quad (3.5)$$

Nyt käyttämällä Laplace-Stieltjesin muunnosta voimme rajata  $\mathbb{E}[e^{sX}]$  seuraavasti:

$$\begin{aligned}\mathbb{E}[e^{sX}] &= \int_0^\infty e^{sx} f(x) dx \\ &\geq \int_{x_k}^\infty e^{sx} f(x) dx \\ &\geq e^{sx_k} \bar{F}(x_k).\end{aligned}$$

Koska yllä mainittu laskelma pätee kaikille  $k$ , seuraa yhtälöstä 3.5, että  $\mathbb{E}[e^{sX}] = \infty$ . Eli väittämästä (i) seura väittäjä (ii).

(ii)  $\Rightarrow$  (iii). Oletetaan nyt, että  $X$  täyttää väittämän (ii), mutta ei väittämää (iii). Tämä tarkoittaa siis, että

$$\liminf_{x \rightarrow \infty} \frac{R(x)}{x} > 0.$$

Tästä seuraa, että on olemassa  $\mu > 0$  ja  $x_0 > 0$  siten, että

$$\frac{R(x)}{x} = \frac{-\ln \bar{F}(x)}{x} \geq \mu \iff \bar{F}(x) \leq e^{-\mu x},$$

kaikilla  $x \geq x_0$ . Valitaan nyt  $s$  siten, että  $0 < s < \mu$ . Tästä saamme edelleen

$$\begin{aligned}M(s) = \mathbb{E}[e^{sX}] &= \int_0^\infty \mathbb{P}(e^{sX} > x) dx \\ &= \int_0^{e^{sx_0}} \mathbb{P}(e^{sX} > x) dx + \int_{e^{sx_0}}^\infty \mathbb{P}(X > \frac{-\ln x}{s}) dx.\end{aligned}$$

Nyt koska  $x \geq e^{sx_0} \iff \frac{\ln x}{s} \geq x_0$  saamme

$$\begin{aligned}M(s) &\leq e^{sx_0} + \int_{e^{sx_0}}^\infty e^{-\mu \frac{\ln x}{s}} dx \\ &= e^{sx_0} + \int_{e^{sx_0}}^\infty x^{-\frac{\mu}{s}} dx.\end{aligned}$$

Koska  $\mu/s > 1$  saamme  $\int_1^\infty x^{-\mu/s} dx < \infty$ , joka puolestaan johtaa siihen, että  $M(s) < \infty$ . Tästä seuraa ristiriita, joten väittämästä (ii) seuraa väittäjä (iii).

(iii)  $\Rightarrow$  (i). Oletetaan vielä lopuksi, että satunnaismuuttuja  $X$  toteuttaa väittämän (iii). Tämän seurauksena on olemassa aidosti kasvava sarja  $(x_k)_{k \leq 1}$  toteuttaen ehdon  $\lim_{k \rightarrow \infty} x_k = \infty$ , siten että

$$\liminf_{x \rightarrow \infty} \frac{R(x_k)}{x_k} = 0.$$

Kun  $\mu > 0$ , tämä ilmaisee että on olemassa  $k_0 \in \mathbb{N}$  siten, että

$$\begin{aligned} \frac{R(x_k)}{x_k} &< e^{-\mu/2} \\ \iff \bar{F}(x_k) &> e^{-\mu x_k/2} \\ \iff \frac{\bar{F}(x_k)}{e^{-\mu x_k}} &> e^{\mu x_k/2}, \end{aligned}$$

kaikilla  $k > k_0$ . Viimeinen edellä oleva väite ilmaisee, että  $\lim_{k \rightarrow \infty} \frac{\bar{F}(x_k)}{e^{-\mu x_k}} = \infty$ , joka puolestaan johtaa siihen, että  $\limsup_{x \rightarrow \infty} \frac{\bar{F}(x)}{e^{-\mu x}} = \infty$ . Koska tämä on totta kaikilla  $\mu > 0$ , saamme että väittämästä (iii) seuraa myös väittämä (i). [22]  $\square$

**Lause 3.2.** *Olkoon kertymäfunktio  $F$  jatkuva tiheysfunktiolla  $f$ . Oletetaan, että  $F$  on paksuhäntäinen. Silloin myös sen tiheysfunktio  $f(x)$  on paksuhäntäinen.*

*Todistus.* Oletetaan, että  $f(x)$  ei ole paksuhäntäinen. Silloin on olemassa  $\lambda' > 0$  ja  $x_0$ , siten että

$$c := \sup_{x > x_0} f(x)e^{\lambda'x} < \infty,$$

ja siksi kaikille  $\lambda \in (0, \lambda')$

$$\int_{\mathbb{R}} e^{\lambda x} f(x) dx \leq e^{\lambda x_0} + c \int_{x_0}^{\infty} e^{\lambda x} e^{-\lambda' x} dx < \infty$$

Tästä seuraa, että kohdassa 3.3 määritelty integraali on äärellinen kaikille  $\lambda$  siten, että  $0 < \lambda < \lambda'$ , mikä on ristiriidassa kertymäfunktion  $F$  paksuhäntäisyyden kanssa.  $\square$

Hilajakauille (lattice distribution) saadaan seuraava tulos.

**Lause 3.3.** *Olkoon  $F$  kertymäfunktio jollekin hilalle  $\{a + hn, n \in \mathbb{Z}\}$ ,  $a \in \mathbb{R}, h > 0$  todennäköisyyksillä  $F\{a + hn\} = p_n$ . Tällöin  $F$  on paksuhäntäinen jos ja vain jos joukko  $\{p_n\}$  on paksuhäntäinen, eli*

$$\limsup_{n \rightarrow \infty} p_n e^{\lambda n} = \infty \quad \text{kaikilla } \lambda > 0. \quad (3.6)$$

*Todistus.* Todistus seuraa lauseesta 3.1 kun  $T = h$ . [12]  $\square$

Tarkastellaan seuraavaksi tarkemmin yhtä yleisesti tunnettua paksuhäntäistä jakaumaa, yleistettyä Pareto-jakaumaa.



### 3.2 Yleistetty Pareto-jakauma

Kynnysarvon ylitys -menetelmäksi kutsutaan menetelmää, joka tarkastelee otoksen arvoja, jotka ylittävät tietyn raja-arvon  $u$ . Raja-arvon  $u$  ylittäneille arvoille voidaan muodostaa kertymäfunktio

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}, \quad (3.7)$$

jossa  $0 \leq x < x_F - u$ . Nyt kun raja-arvon  $u$  suuruus lähenee otoksen päätepistettä  $x_F$  voidaan ylitysten jakaumaa approksimoida yleistetyllä Pareto-jakaumalla. [3] Tapahtumaa voidaan kuvata matemaattisesti

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi, \sigma(u)}(x)| = 0, \quad (3.8)$$

missä  $G_{\xi, \sigma(u)}$  kuvaa yleistetyn Pareto-jakauman kertymäfunktioita (määritellään alla),  $\sigma(u)$  kuvaa skaalaparametriä ja  $\xi$  on jakauman muotoparametri [24].

Skaalaparametri  $\sigma(u)$  on raja-arvon  $u$  funktio, sillä sitä arvioidaan ylityksien  $y_i$  kautta. Siihen vaikuttavat siis ylityksien kappalemäärä sekä niiden sijainti. Muotoparametri  $\xi$  puolestaan kertoo jakauman hännän vähenemisnopeutta, joten se ei ole riippuvainen valitusta raja-arvosta, kunhan otoksessa on sen ylittäviä arvoja riittävästi. Määritellään seuraavaksi vielä yleistetyn Pareto-jakauman kertymäfunktio.

**Määritelmä 3.8.** Yleistetyn Pareto-jakauman kertymäfunktio voidaan määritellä

$$G_{\xi, \sigma(u)}(x) = \begin{cases} 1 - (1 + \frac{\xi x}{\sigma(u)})^{-1/\xi}, & \text{kun } \xi \neq 0. \\ 1 - e^{-\frac{x}{\sigma(u)}}, & \text{kun } \xi = 0, \end{cases} \quad (3.9)$$

missä

$$\begin{aligned} x &\geq 0, & \text{jos } \xi &\geq 0, \\ 0 \leq x &\leq -\frac{\sigma(u)}{\xi}, & \text{kun } \xi &< 0 \end{aligned}$$

ja skaalaparametri  $\sigma(u) > 0$ . [11]

Vastaavat tiheysfunktiot ovat

$$g_{\xi, \sigma(u)}(x) = \begin{cases} \frac{1}{\sigma(u)} (1 + \frac{\xi x}{\sigma(u)})^{-(\frac{1}{\xi} + 1)}, & \text{kun } \xi \neq 0. \\ \frac{1}{\sigma(u)} e^{-\frac{x}{\sigma(u)}}, & \text{kun } \xi = 0. \end{cases} \quad (3.10)$$

Määritelmässä 3.8 määritellyn yleistetyn Pareto-jakauman kertymäfunktion  $G_{\xi, \sigma(u)}$  rajarvoksi voidaan tulkita tilanne  $\xi = 0(G_{0, \sigma(u)})$ , kun  $\xi \rightarrow 0$ . Tästä saadaan [11]

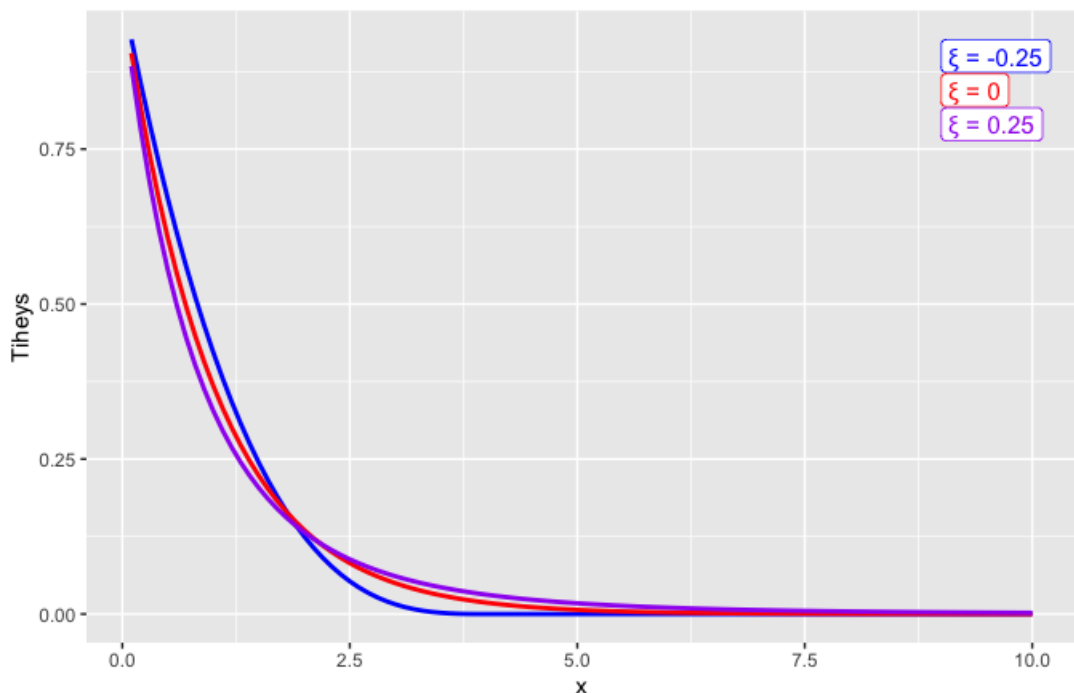
$$G_{\xi, \sigma(u)}(x) = 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-1/\xi}, \quad x \in D(\xi, \sigma), \quad (3.11)$$

missä

$$x \in D(\xi, \sigma) = \begin{cases} [0, \infty], & \text{kun } \xi \geq 0, \\ [0, -\sigma/\xi], & \text{kun } \xi < 0. \end{cases}$$

Yleistetty Pareto-jakauma, kuten nimikin jo paljastaa, on yleistetty esitysmuoto. Esimerkkejä yleistetyn Pareto-jakauman muodoista ovat tasajakauma ( $\xi = -1$ ), eksponenttijakauma ( $\xi = 0$ ) ja Pareto-jakauma ( $\xi > 0$ ).

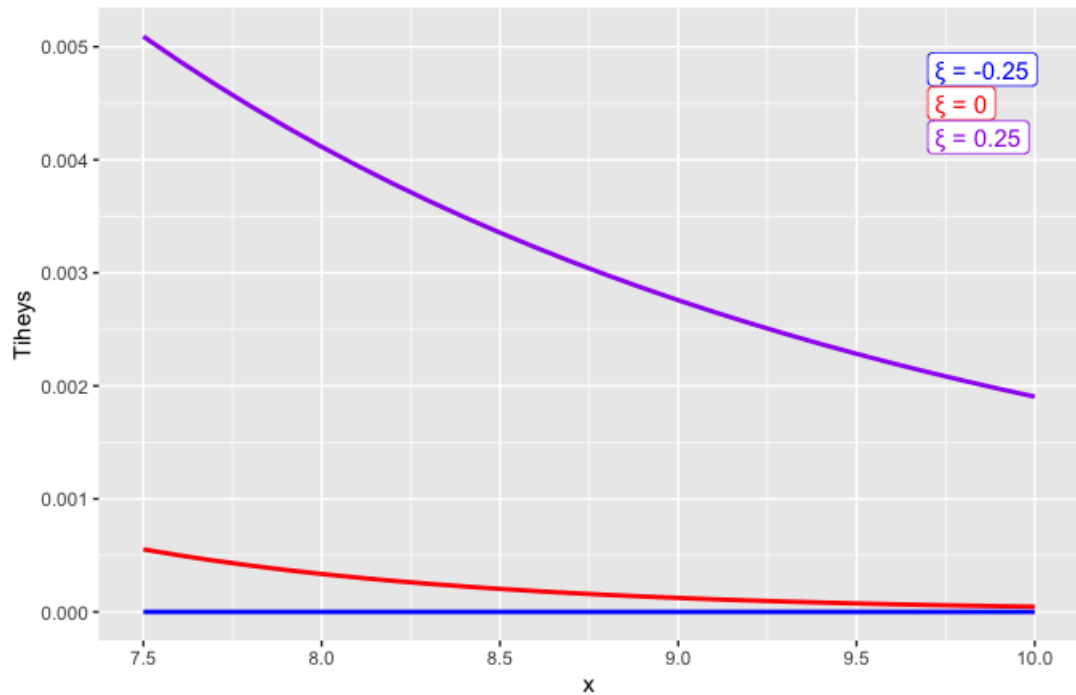
Tarkastellaan seuraavaksi yleistetyn Pareto-jakauman tiheysfunktioita eri muotoparametreillä  $\xi = -0.25$ ,  $\xi = 0$  ja  $\xi = 0.25$  skaalaparamterin  $\sigma(u)$  pysyessä vakiona. Nämä on esitetty kuvassa 3.1 alla



**Kuva 3.1.** Yleistetyn Pareto-jakauman tiheysfunktioita eri muotoparametreillä (mukaillen [11])

Kuvasta 3.1 huomataan, että muotoparametrin kasvaessa kuvaaja laskee jyrkemmin, mutta laskun jälkeen se loivenee nopeammin ja sen häntä on paksumpi. Koska olem-

me tässä työssä kiinnostuneita erityisesti jakaumien häntäarvoista ja kuvasta 3.1 on hie-  
man hankala päätellä kuvaajan hännän käyttäytymistä, piirretään vielä toinen kuva 3.2,  
jossa tiheysfunktioiden kuvaajat on esitetty välillä 7.5-10.



**Kuva 3.2.** Yleistetyn Pareto-jakauman tiheysfunktioiden arvot välillä 7.5 - 10 eri muoto-  
parametreilla

Kuvasta 3.2 huomaamme, että päätelmämme aikaisemmasta kuvasta olivat oikeat eli  
muoto parametrin kasvattaminen kasvattaa myös yleistetyn Pareto-jakauman hännän pak-  
suutta. Kuten kuvista 3.1 ja 3.2, sekä paksuhäntäisen jakauman määritelmästä nähdään,  
niin Pareto-jakaumaa (lila käyrä) voidaan pitää määritelmän mukaan paksuhäntäisenä,  
sillä sen häntä on paksumpi kuin eksponenttijakauman (punainen käyrä).

### 3.3 Muita paksuhäntäisiä jakaumia

Kuten aikaisemmin jo todettiin yleistetyn Pareto-jakauman lisäksi on olemassa myös mui-  
ta paksuhäntäisiä jakaumia, jotka voivat olla jossain tilanteissa sopivampia kuin yleistetty  
Pareto-jakauma. Tarkastellaan nyt vielä tarkemmin jo mainittujen Cauchyn-jakauman ja  
Burr-jakauman ominaisuuksia yleisellä tasolla.

### 3.3.1 Cauchy-jakauma

Cauchyn jakauman tiheysfunktion yleinen kaava on

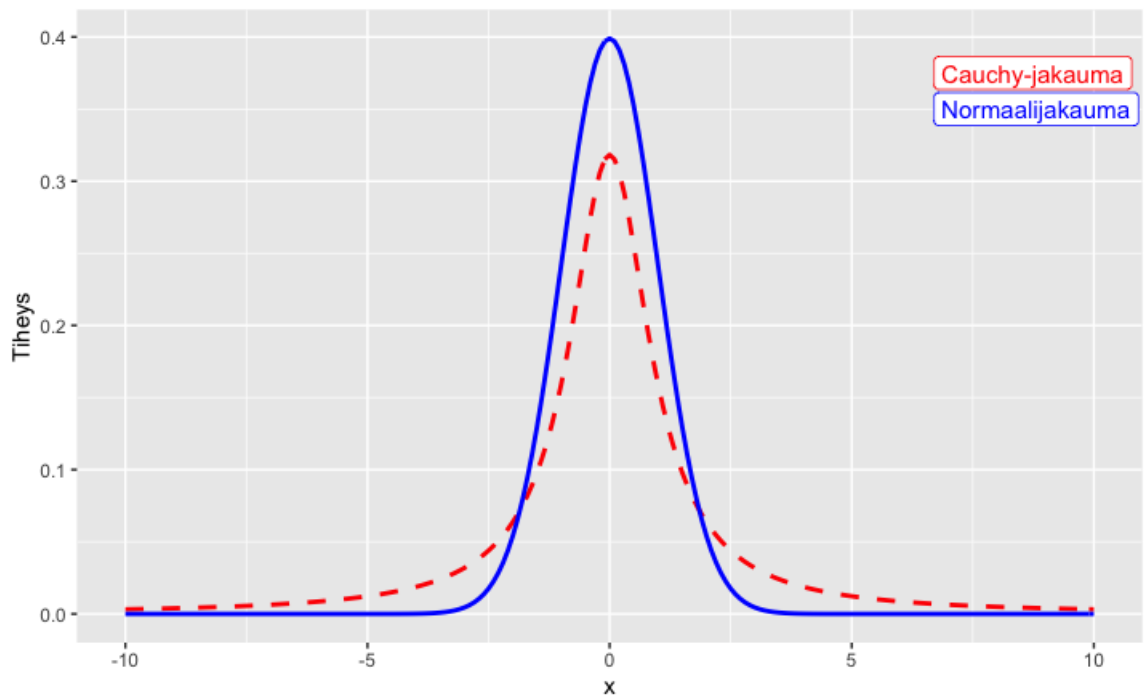
$$f(x) = \frac{1}{s\pi(1 + ((x - t)/s)^2)}, \quad (3.12)$$

missä  $t$  on sijaintiparametri ja  $s$  on mittakaavaparametri. Tapausta, jossa  $t = 0$  ja  $s = 1$  kutsutaan Cauchyn standardijakaumaksi. Standardin Cauchyn jakauman yhtälö supistuu muotoon

$$f(x) = \frac{1}{\pi(1 + x^2)}. \quad (3.13)$$

Cauchyn jakaumat muistuttavat normaalijakaumaa muotonsa puolesta, mutta niillä on kuitenkin selkeästi paksummat hännät. Yleisesti ottaen ne soveltuvatkin hyvin kun tutkitaan hypoteesejä, joiden oletuksena on normalisuus. Cauchyn jakauman erikoisuutena on sen keskiarvon ja keskihajonnan puuttuminen / määrittelemättömyys.

Cauchyn standardijakauman tiheysfunktion kuvaajan muotoa on havainnollistettu kuvassa 3.3 vertaamalla sitä normaalijakauman tiheysfunktion kuvaajaan. Cauchy-jakauman sijaintiparametriksi on määritely 0 ja muotoparametriksi 1, jotta se sijoittuisi mahdollisimman hyvin standardin normaalijakauman kohdalle. [10]



**Kuva 3.3.** Normaalijakauman ja Cauchy-jakauman tiheysfunktiot (mukaan [19])

Cauchyn jakauman kertymäfunktio on muotoa [10]

$$F(x) = 0.5 + \frac{\arctan(x)}{\pi}. \quad (3.14)$$

### 3.3.2 Burr-jakauma

Burr-jakauma on Irving Wingate Burrin vuonna 1942 ensimmäisen kerran esittelemä jakauma. Se soveltuu erityisesti oikealla vinouteen paksuhäntäisen datan analysointiin. Burr-jakaumalla on kaksi positiivista parametriä, muotoparametri  $k$  ja aste parametri  $c$ . Tämä johtaa todennäköisyysfunktioon, joka voi olla joko laskeva tai unimodaalinen. Lisäksi sen hazard rate- funktio voi olla joko laskeva tai ylösalaisin oleva amme. Burr-jakaumaa käytetään laajasti eloonjäämis- sekä parantumisajan mallintamiseen lääketieteen alalla sekä kotitalouksien tulojen mallintamisessa. [15]

Burrin vuonna 1942 esittelemä differentiaaliyhtälö oli

$$\frac{dF(x)}{dx} = F(x)[1 - F(x)]g(x), \quad (3.15)$$

jossa  $0 \leq F(x) \leq 1$  ja funktio  $g(x)$  on jatkuva ja positiivinen. Yhtälö saa erityyppisiä ominaisuuksia riippuen funktion  $g(x)$  valinnoista, joten Burr määrittä 12 jatkuvan jakauman tyyppiä ja muodosti niistä niin kutsutun Burrin järjestelmän. Seuraavalla sivulla olevassa kuvassa on esitelty nämä erilaiset jakaumatyyppit.

Type	$F(x)$	Interval
I	$x$	$(0,1)$
II	$(1 + e^{-x})^{-k}$	$(-\infty, \infty)$
III	$(1 + x^{-c})^{-k}$	$(0, \infty)$
IV	$\{1 + [x^{-1}(c - x)]^{1/c}\}^{-k}$	$(0, c)$
V	$[1 + ce^{-\tan(x)}]^{-k}$	$(-\pi/2, \pi/2)$
VI	$[1 + ce^{-r \sinh(x)}]^{-k}$	$(-\infty, \infty)$
VII	$2^{-k} [1 + \tanh(x)]^k$	$(-\infty, \infty)$
VIII	$[2\pi^{-1} \arctan(e^x)]^{-k}$	$(-\infty, \infty)$
IX	$1 - 2\{2 + c[(1 + e^x)^k - 1]\}^{-1}$	$(-\infty, \infty)$
X	$(1 - e^{-x^2})^k$	$(0, \infty)$
XI	$[x - (2\pi)^{-1} \sin(2\pi x)]^k$	$(0, 1)$
XII	$1 - (1 + x^c)^{-k}$	$(0, \infty)$

**Kuva 3.4.** BURR-jakauman 12 eri tyyppiä [15]

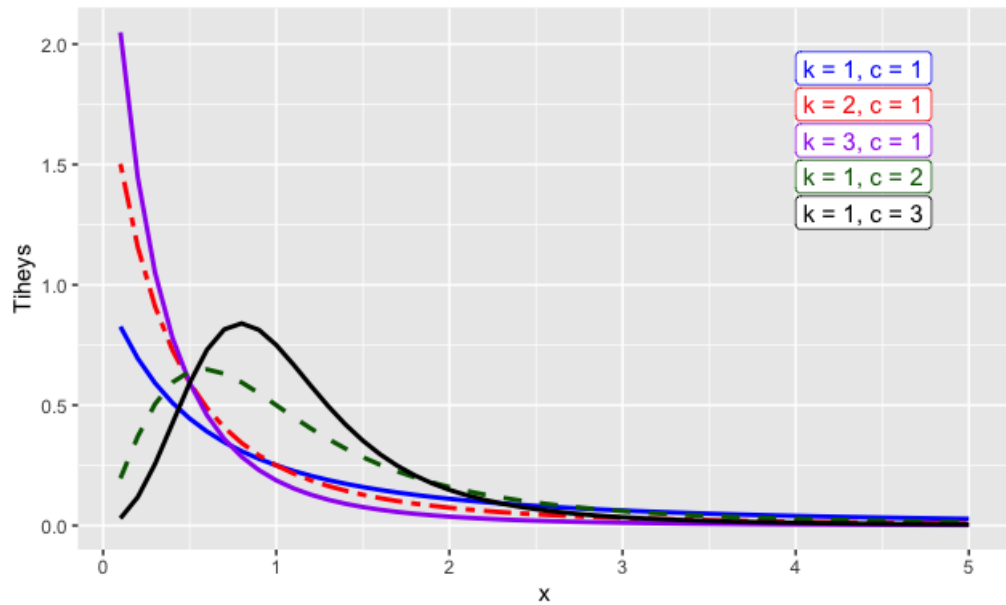
Burrin järjestelmän tärkein jakauma on Burrin tyyppi XII (12), joka onkin jakaumista tunnetuin ja käytetyin. Sen tunnettavuuden vuoksi Burrin tyyppin XII jakaumaa onkin usein kutsuttu Burr-jakaumaksi. Kuvasta (3.4) nähdään, että Burr-jakauman satunnaismuuttujan  $X$  kertymäfunktio  $F$  määritellään seuraavasti

$$F(x) = P(X \leq x) = 1 - (1 + x^c)^{-k}, \quad x \geq 0. \quad (3.16)$$

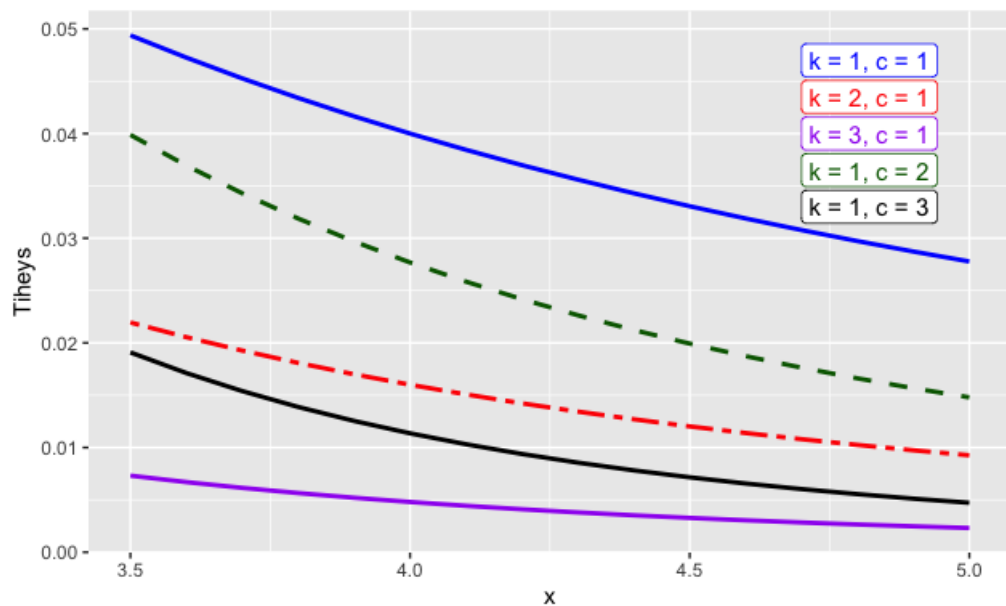
Yhtälöstä (3.15) saadaan nyt myös Burr-jakauman tiheysfunktio, joka on muotoa [15]

$$f(x) = \frac{dF(x)}{dx} = kcx^{c-1}(1 + x^c)^{-(k+1)}, \quad x \geq 0. \quad (3.17)$$

Kuvissa 3.5 ja 3.6 on vielä havainnollistettu, kuinka muoto- ja asteparametrien valinta vaikuttaa Burr-jakauman tiheysfunktion kuvaajaan.



**Kuva 3.5.** Burr-jakauman tiheysfunktioita eri muoto- ja asteparametreilla. (mukailten [15] ja [17])



**Kuva 3.6.** Burr-jakauman tiheysfunktioiden kuvaajat välillä 3.5 - 5 eri muoto- ja asteparametreilla.

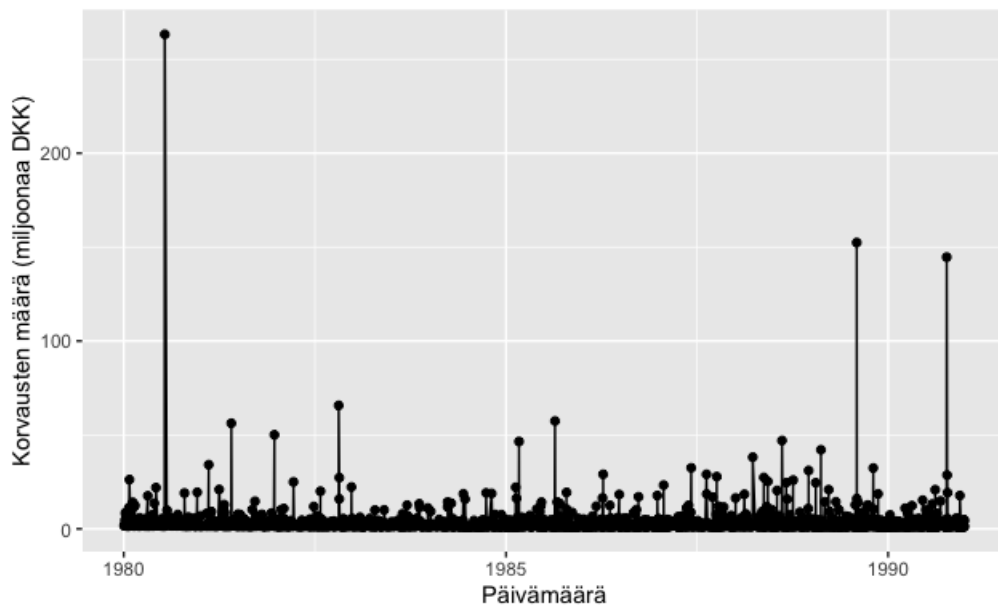
Kuvista 3.5 ja 3.6 huomataan, että muotoparametrin kasvattaminen johtaa Burr-jakauman vähenemisnopeuden kasvamiseen ja hännän ohentamiseen. Asteparametrin kasvaminen puolestaan johtaa jakauman huipun siirtymiseen oikealle ja ohentaa myös jakauman häntää.

## 4. YLEISTETYN PARETO-JAKAUMAN KÄYTTÖ MALLINTAMISESSA

Tässä kappaleessa testataan R-ohjelmiston avulla yleistetyn Pareto-jakauman sopivuutta käytännön dataan ja sovitetaan se dataan estimoimalla sille oikeat parametrit. Lopuksi estimoinnin tulosten pohjalta datalle lasketaan vielä odotettujen tappioiden arvot muuttamalla eri todennäköisyydellä.

### 4.1 Tanskalainen tulipalodata

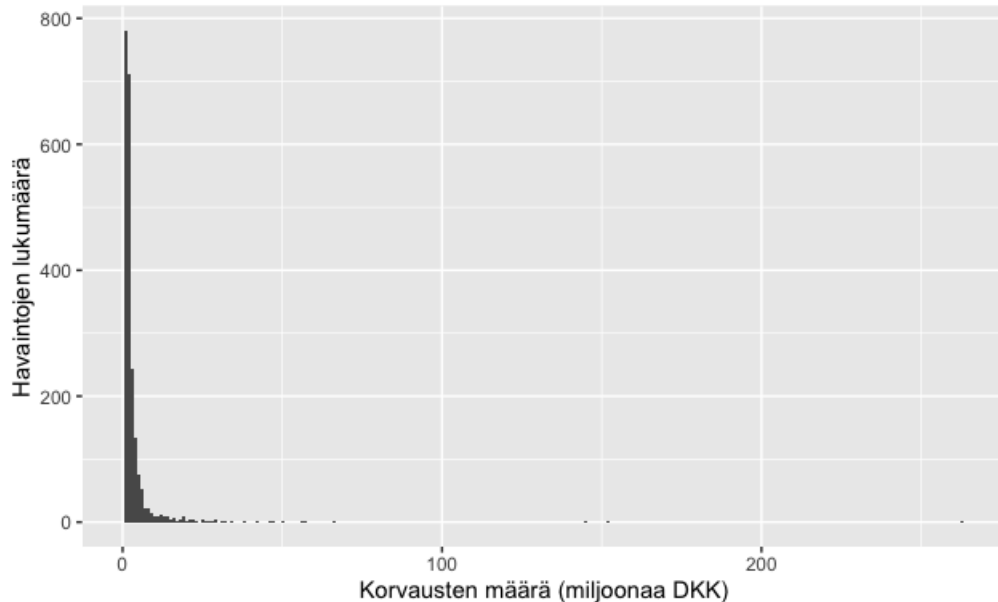
R-ohjelmistossa on ladattavissa data, joka sisältää Tanskassa sattuneiden tulipalojen johdosta maksettavien palovakuutuskorvausten suuruuksia. Datan on toimittanut Mette Rytgaard ja se on kerätty torstain 3. tammikuuta 1980 ja maanantain 31. joulukuuta 1990 väliseltä ajalta. Tiedoston lataamiseksi R-ohjelmistossa on oltava ladattuna "evir"-paketti. [9] Data sisältää yhteensä 2167 tapahtumaa eri ajankohdilta ja niistä koituneet korvaukset on havainnollistettu kuvassa 4.1 alla.



**Kuva 4.1.** Tanskan data tulipalovakuutuskorvausten suuruuksista aikavälillä 1980 - 1990.



Kuvasta 4.1 huomataan, että suurin osa korvauksista on alle 30 miljoonaa Tanskan kruunua (DKK), mutta muutaman päivän kohdalla korvaukset ovat huomattavasti muita arvoja suurempia. Piirretään vielä datasta histogrammi, jotta nähdään paremmin koituneiden korvausten jakauma.



**Kuva 4.2.** Histogrammi Tanskan tulipalodatasta

Kuva 4.2 tukee aikaisempaa havaintoamme siitä, että suurin osa vahingoista aiheutuneista korvauksista on alle 30 miljoonaa DKK. Suuremmat yksittäiset tapahtumat on havaittavissa yksittäisinä pisteinä esimerkiksi noin 150:n ja 270:n miljoonan DKK:n kohdalla. Kyseinen data on oppikirjaesimerkki tilanteesta, jonka mallinnuksessa kannattaa hyödyntää paksuhäntäisiä jakaumia normaalijakauman sijasta. Mikäli tutkisimme dataa normaalijakauman avulla, yksittäiset suuret korvaukset saisivat esiintymistodennäköisyydekseen nolla, vaikka niiden sattuessa kohdalle vaikutus on todella merkittävä. Tästä syystä suuret vahingot on ehdottomasti otettava huomioon myös mallinnuksessa.

## 4.2 Datan mallinnus yleistetyllä Pareto-jakaumalla

Yleistetyn Pareto-jakauman parametrit on estimoitava, jotta jakauma saadaan mahdollisimman hyvin mallinnettua havaintoihin sopiviksi. On myös syytä suorittaa yleinen mallin sopivuuden tarkastelu. Tässä kappaleessa käsitellään ensin mallin ja parametrien estimointia yleisellä tasolla ja suoritetaan sen jälkeen mallinnus hyödyntäen R-ohjelmistoa.

### 4.2.1 Mallin hyvyyden ja parametrien estimointi

Yleistetty Pareto-jakauma sisältää 3 merkittävää parametriä: raja-arvon ( $u$ ), skaalaparametrin ( $\sigma$ ) ja muotoparametrin ( $\xi$ ). Yksi yleistetyn Pareto-jakauman sopivuuden edellytyksistä on oikea raja-arvon valinta. Raja-arvo täytyy valita siten, että se on tarpeeksi korkea, jotta ylityksiä ei tule liikaa, mutta niitä tulee kuitenkin olla tarpeeksi, jotta niiden muodostama funktio on pätevä. Merkitään raja-arvon ylittävien tapausten määrää  $N_u$ :lla ja ylitysten suuruuksia  $Y_i = X_i - u$ , kun  $X_i > u$ . Nyt, mikäli raja-arvo on valittu hyvin, voidaan ylitysten ( $Y_i$ ) muodostamaa jakaumaa  $F_u$  kuvata yhtälöiden (3.7) ja (3.8) mukaan yleistetyllä Pareto-jakaumalla ja esittää se muodossa

$$G_{\hat{\xi}, \hat{\sigma}}(y) \approx \hat{F}_u(y) = \frac{\hat{F}(y+u) - \hat{F}(u)}{1 - \hat{F}(u)}, \quad (4.1)$$

missä  $\hat{F}$  on datan havaintojen empiirinen kertymäfunktio ja "hattu"parametrien päällä kuvastaa estimaattia.

Kun data on tarpeeksi suuri voidaan yhtälössä 4.1 esiintyvää havaintojen empiiristä kertymäfunktioita pisteessä  $u$ , kuvata  $\hat{F}(u) = 1 - \frac{N_u}{n}$ , missä  $N_u$  kuvaa raja-arvon  $u$  ylittävien havaintojen kappalemäärää ja  $n$  puolestaan kaikkien havaintojen kappalemäärää. Nyt kaava (4.1) voidaan kirjoittaa muotoon

$$G_{\hat{\xi}, \hat{\sigma}}(y) = \frac{\hat{F}(y+u) - [1 - \frac{N_u}{n}]}{\frac{N_u}{n}}, \quad (4.2)$$

josta voidaan edelleen ratkaista alkuperäisten havaintojen  $X_i = Y_i + u$  kertymäfunktion hännän kaava hyödyntämällä yhtälöä (3.11)

$$\begin{aligned} \hat{F}(y+u) &= \frac{N_u}{n} G_{\hat{\xi}, \hat{\sigma}}(y) + [1 - \frac{N_u}{n}] \\ &= 1 - \frac{N_u}{n} \left(1 - G_{\hat{\xi}, \hat{\sigma}}(y)\right) \\ &= 1 - \frac{N_u}{n} \left(1 + \hat{\xi} \frac{y}{\hat{\sigma}}\right)^{-1/\hat{\xi}} \\ \hat{F}(x) &= 1 - \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x-u}{\hat{\sigma}}\right)^{-1/\hat{\xi}}. \end{aligned} \quad (4.3)$$

[28] Kääntämällä (4.3) saadaan vielä ratkaistua  $\hat{F}(x)$ :n  $q$ :nnes kvantiiliestimaatti  $\hat{x}_q$

$$\hat{x}_q = \hat{F}^{-1}(q) = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[ \left( \frac{n}{N_u} (1-q) \right)^{-\hat{\xi}} - 1 \right], \quad (4.4)$$

jota voidaan käyttää yleistetyn Pareto-jakauman hännän sopivuuden arvioinnissa suh-

teessa dataan [11].

Raja-arvon  $u$  lisäksi meidän on estimoitava myös skaala- ( $\xi$ ) ja muotoparametrit ( $\sigma$ ). Näiden parametrien estimoinnissa voidaan käyttää useita eri menetelmiä, mutta käydään tässä läpi menetelmistä yleisin, MLE (maximum likelihood estimation) eli suurimman uskottavuuden estimaatti.

Käytetään estimointiin aineistoa, jossa toisistaan riippumattomia havaintoja kuvataan  $X$ :llä ( $X_1, \dots, X_n$ ), valitun raja-arvon  $u$  ylittäviä havaintoja  $Y$ :llä ( $Y_i = X_i - u, X_i > u$ ) ja yleistetyn Pareto-jakauman kertymäfunktioita, parametreilla  $\xi$  ja  $\sigma$ ,  $G_{\xi,\sigma}$ :llä. Täten vastaava tiheysfunktio saadaan yhtälöstä (3.10)

$$g_{\xi,\sigma}(y) = G'_{\xi,\sigma}(y) = \frac{1}{\sigma} \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi-1}, \quad y \in D(\xi, \sigma). \quad (4.5)$$

[28] Uskottavuusfunktio voidaan puolestaan ilmaista muodossa

$$L(\xi, \sigma; Y) = \prod_{i=1}^n g_{\xi,\sigma}(Y_i), \quad (4.6)$$

missä  $1 + \xi \frac{Y_i}{\sigma} > 0$ . Muodostetaan uskottavuusfunktion pohjalta vielä log-uskottavuusfunktio

$$\begin{aligned} l(\xi, \sigma; Y) &= \sum_{i=1}^n \ln \left[ \frac{1}{\sigma} \left(1 + \xi \frac{Y_i}{\sigma}\right)^{-1/\xi-1} \right] \\ &= -n \ln \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \log \left(1 + \frac{\xi}{\sigma} Y_i\right), \end{aligned} \quad (4.7)$$

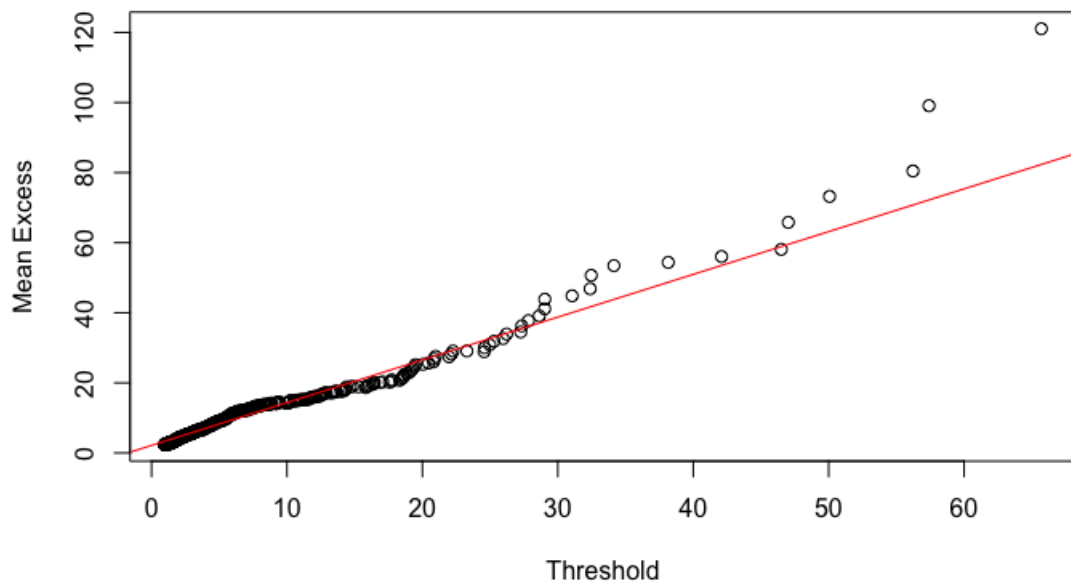
josta parametrit pystytään nyt ratkaisemaan maksimoimalla kyseinen funktio [23]

$$\begin{cases} \frac{\partial l}{\partial \xi} = 0 \\ \frac{\partial l}{\partial \sigma} = 0 \end{cases} \Rightarrow \begin{cases} n &= (\hat{\xi} - 1) \sum_{i=1}^n \left(1 - \frac{\hat{\xi}}{\sigma} Y_i\right)^{-1} \\ \hat{\xi} &= -\frac{1}{n} \sum_{i=1}^n \log \left(1 - \frac{\hat{\xi}}{\sigma} Y_i\right). \end{cases} \quad (4.8)$$

## 4.2.2 Mallin sovitus dataan

Pyritään seuraavaksi mallintamaan kappaleessa 4.1 esitelty tanskalainen tulipalodata hyödyntäen yleistettyä Pareto-jakaumaa ja R-ohjelmistoa. Aloitetaan mallinnus tarkastamalla käyttäytykö tarkastelemamme data paksuhäntäisesti. Suoritetaan datan paksuhäntäisyyden testaus piirtämällä sille keskimääräisen ylityksen funktio ("mean excess

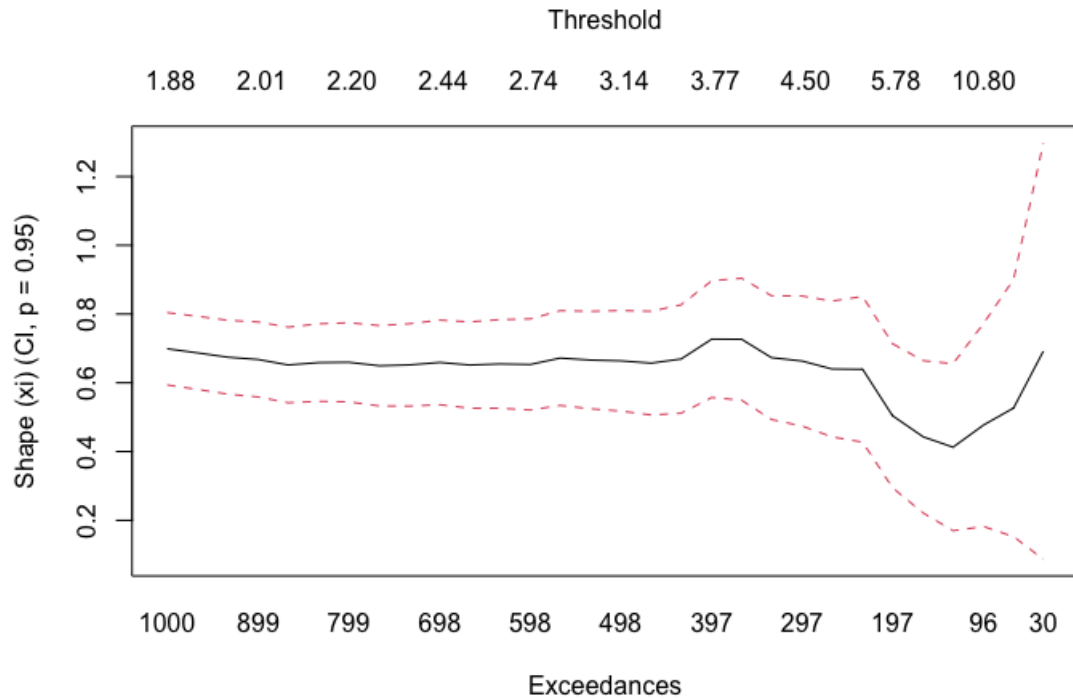
function"). Keskimääräisen ylityksen funktion kuvaajan ollessa nouseva datan voidaan katsoa olevan paksuhäntäistä. Mikäli kuvaaja puolestaan olisi laskeva, datan katsottaisiin olevan ohuthäntäisesti käyttäytyvä. Kulmakertoimella nolla eli vaakasuora kuvaaja puolestaan ilmaisisi datan olevan eksponentisesti jakautunutta. [14]



**Kuva 4.3.** Keskimääräisen ylityksen funktion -kuvaaja tanskalaiselle tulipalodatalle.

Yllä olevasta kuvasta 4.3 huomataan, että keskimääräisen ylityksen funktion kuvaajan käyrä on nouseva eli dataa voidaan pitää paksuhäntäisesti jakautuneena. Nyt kun paksuhäntäisyys on varmistettu voidaan siirtyä itse mallintamiseen.

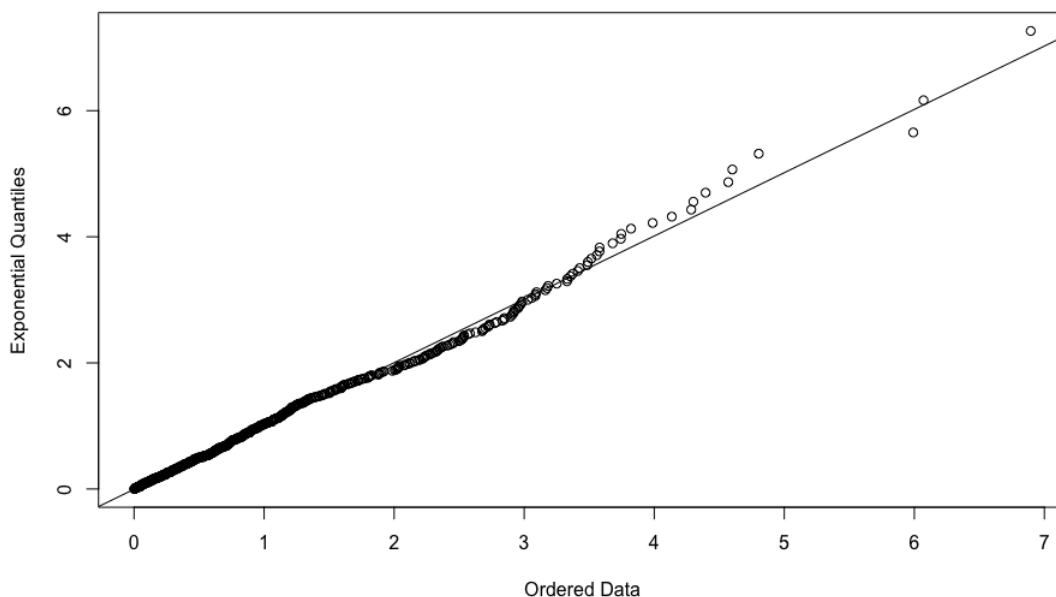
Kuten kappaleesta 4.2.1 huomataan, täytyy mallinnus aloittaa parametrien estimoinnilla ja parametrien estimointi puolestaan sopivan raja-arvon valinnalla. Paras mahdollinen raja-arvo voidaan laskea esimerkiksi haarukointi-tekniikalla. Käytännössä dataan siis sovitetaan useampi yleistetty pareto-jakauma eri raja-arvoilla ja pyritään löytämään se arvo, jolla sovitettu jakauma poikkeaa varianssin osalta mahdollisimman vähän toteutuneista arvoista. Tällä tavalla laskettaessa tanskalaiselle tulipalodatalle saadaan estimoitua raja-arvoksi noin 2,4 miljoonaa Tanskan kruunua. R-ohjelmistosta löytyy myös komento "shape", jonka avulla oikean raja-arvon valintaa voidaan analysoida visuaalisesti. Kuvassa seuraavalla sivulla on esitetty "shape"komennon tuottama kuvaaja tanskalaiselle tulipalodatalle.



**Kuva 4.4.** Raja-arvon valintaa helpottavan "shape"-komennon tuottama kuvaaja R-ohjelmistossa

Kuvasta 4.4 huomataan, että yllä lasketulla raja-arvolla 2,4 miljoonaa Tanskan kruunua, raja-arvon ylittäneitä tapauksia on noin 700 kappaletta ja yleistetty Pareto-jakauma sopii dataan parhaiten muotoparametrin ollessa suurinpiirtein 0.65. Kuvan 4.4 mukaan myös hieman suurempi raja-arvo voisi sopia tarkasteluun, sillä esimerkiksi raja-arvolla 3, muotoparametrin arvo ei poikkea merkittävästi pienempien raja-arvojen estimoiduista muotoparametrien arvoista. Hyväksytään kuitenkin aiemmin laskettu raja-arvo 2,4 jatkotarkastelua varten, sillä sen avulla raja-arvon ylittäviä tapauksia on varmasti tarpeeksi ja muotoparametrin kuvaaja on hieman vakaampi (pienempi varianssi) kuin suuremmilla raja-arvoilla.

Seuraavaksi arvioidaan yleistetyn Pareto-jakauman sopivuutta dataan estimoimalla sille muoto- ja skaalaparametrit R-ohjelmiston avulla. Tämän jälkeen jakauman sopivuutta dataan testataan kvantiilikuvion (QQ-kuvaaja) avulla. QQ-kuvaaja kuvaa kuinka testatun jakauman tuottamat arvot vastaavat aineistossa esiintyviä havaintoja. Mikäli malli sopii aineiston kuvaamiseen ja sen parametrit ovat oikein määritellyt QQ-kuvaajan pisteet sijaitsevat suoralla. [20] Suoritetaan parametrien esitointi suurimman uskottavuuden menetelmällä hyödyntäen R-ohjelmistoa ja tulostetaan saatu QQ-kuvaaja.



**Kuva 4.5.** Yleistetyin Pareto-jakauman sopivuuden tarkastelu Tanskan tulipalodataan hyödyntäen QQ-kuvaajaa

Kuvasta 4.5 huomataan, että parametrien estimointi on hyvä ja QQ-kuvaaja on likimain suora. R-ohjelmisto estimoi MLE:llä parameterien arvoiksi  $\xi = 0.662$  ja  $\sigma = 1.82$  yllä estimoidulla raja-arvolla  $u = 2.4$ . (katso laskennassa käytetty koodi liitteen kuvasta A.2). Tämä tarkoittaa siis sitä, että yleistetty Pareto-jakauma sopii estimoiduilla parametreilla kuvaamaan aineiston havaintoja.

### 4.2.3 VaR- ja ES-lukujen laskeminen

Lasketaan seuraavaksi vielä edellisessä kappaleessa estimoitujen parametrien pohjalta VaR (Value at risk) ja ES (Expected shortfall, joskus myös cVaR = conditional Value at Risk) tanskalaiselle tulipalodataalle. Nämä molemmat ovat riskienhallinnassa tunnettuja ja laajasti käytettyjä riskin mittareita.

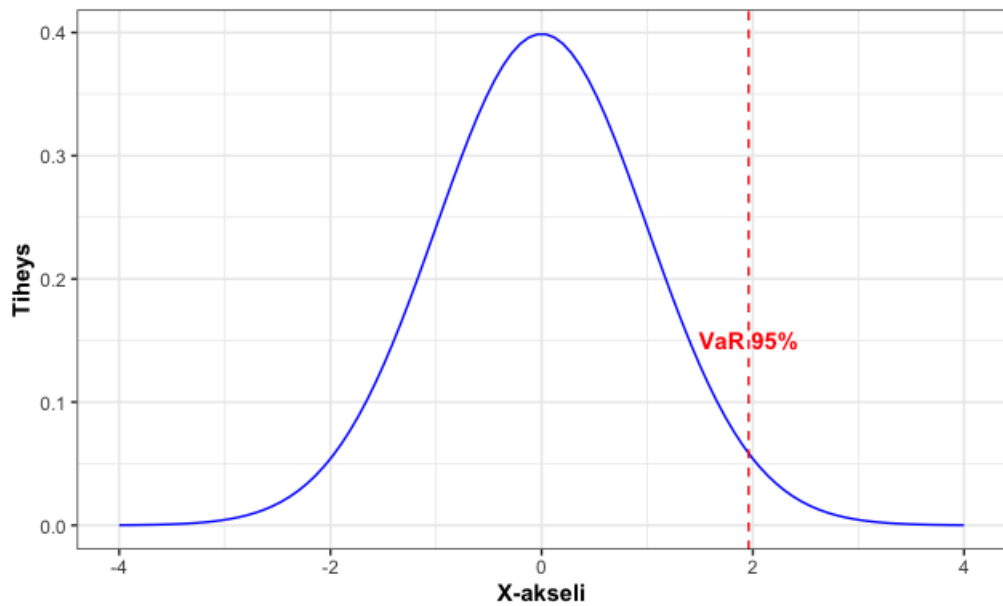
VaR kertoo todennäköisyyden tietynsuuruiseksi tappiolle (tai vastaavasti kuinka suuri tappio tapahtuu tietyllä todennäköisyydellä). VaR:ia käytetään laajasti esimerkiksi rahoitusyriyten ja pankkien keskuudessa. Se soveltuu esimerkiksi sijoitusanalyysiin ja riskialtituden mittaamiseen ja hallintaan. [16]

**Määritelmä 4.1.** Kun  $x \in \mathbb{R}$  ja  $\alpha \in (0, 1)$ , niin VaR:in määritelmäksi saadaan:

$$VaR_\alpha = \inf\{x \mid \mathbb{P}(X \leq x) \geq \alpha\},$$

jossa  $\alpha$  kuvaa haluttua luottamustasoa [2].

VaR on havainnollistettu vielä kuvassa 4.6, jossa on käytetty selkeyden vuoksi normaali-jakaumaa odotusarvolla 0 ja keskihajonnalla 1.



**Kuva 4.6.** 95% luottamustason VaR-arvo normaalijakaumasta, jonka odotusarvo on 0 ja keskihajonta 1. (mukaan [16])

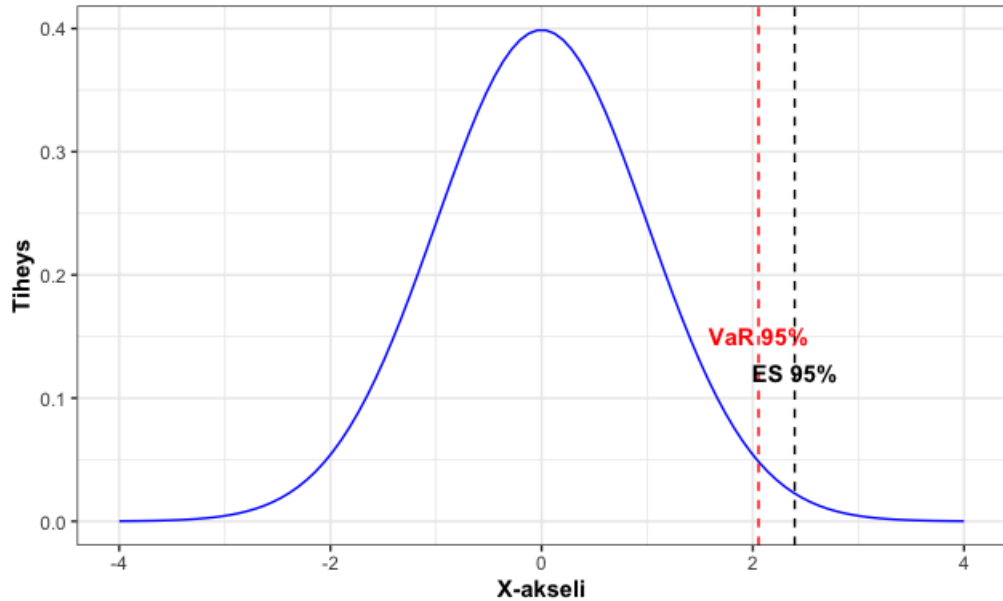
ES puolestaan kertoo odotetun tappion, mikäli VaR-luottamustasoa suurempi tappio tapahtuu. ES on kehitetty siksi, että pelkkä VaR ei ota kantaa tappion suuruuteen mikäli VaR taso ylitetään. [5]

ES määritelläänkin siis yleensä VaR:n avulla:

**Määritelmä 4.2.** Kun  $x \in \mathbb{R}$  ja  $\alpha \in (0, 1)$ , niin ES:n määritelmäksi saadaan:

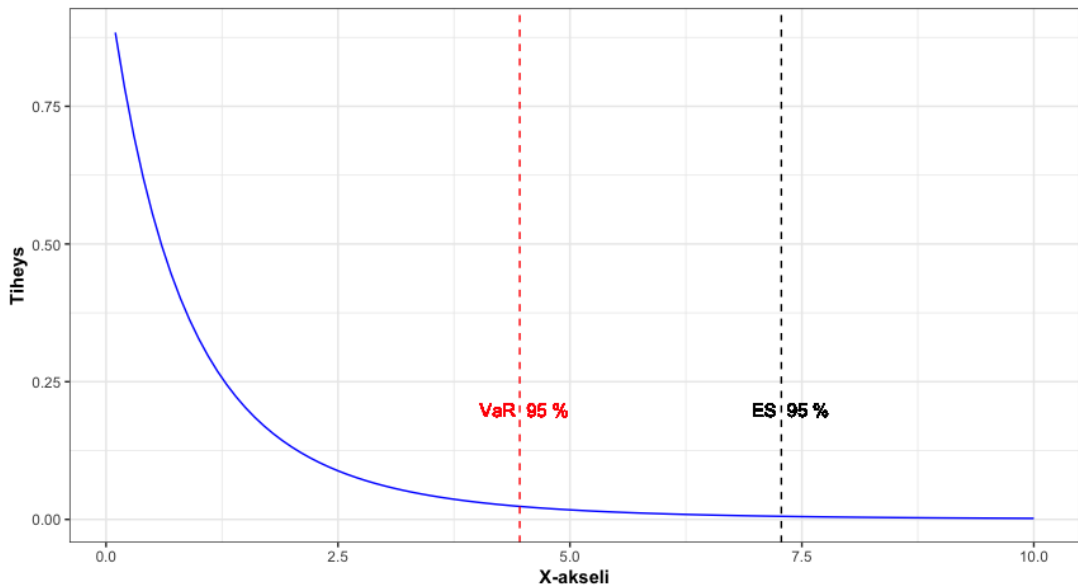
$$\begin{aligned} ES_\alpha &= \frac{1}{1-\alpha} \int_\alpha^1 VaR_x(X) dx \\ &= \mathbb{E}[X \mid X \geq VaR_\alpha(X)] \quad [5]. \end{aligned}$$

Havainnollistetaan vielä myös ES arvoa piirtämällä se samaan kuvaan kuin VaR yllä (4.6):



**Kuva 4.7.** 95% luottamustason VaR- ja ES-arvot normaalijakaumasta, jonka odotusarvo on 0 ja keskihajonta 1. (mukaillen [25])

Kuvasta 4.7 huomataan, että normaalijakauman avulla havainnollistettuna ES ja VaR sijaitsevat melko lähellä toisiaan. Tämä johtuu siitä, että jakauman häntä on melko lyhyt. Piirretään tämän takia vielä vastaavat arvot yleistetylle Pareto-jakaumalle. Käytetään jakauman parametreinä samoja kuin kuvan 3.2 paksuhäntäisimmässä kuvaajassa.



**Kuva 4.8.** 95% luottamustason VaR- ja ES-arvot yleistetylle Pareto-jakaumalle

Kuten kuvasta 4.8 huomaamme, paksuhäntäisen jakauman kohdalla näiden kahden arvot voivat poiketa melko suurestikin toisistaan.



Yleistettyyn Pareto-jakaumaan yhdistettynä VaR:n kaavaksi saadaan

$$VaR = u + \frac{\sigma}{\xi} \left( \left( \frac{n}{N_u} (1-p) \right)^{-\xi} - 1 \right), \quad (4.9)$$

ja ES:n

$$ES = \frac{VaR}{1-\xi} + \frac{\sigma - \xi u}{1-\xi}, \quad (4.10)$$

joissa  $p$  kuvaa haluttua todennäköisyytasoja (esim. 5% todennäköisyydellä tapahtuvaa havaintoa tarkasteltaessa  $p$ :n paikalle asetetaan 0.95 (kuten yllä kuvissa)) ja muut käytetyt parametrit ovat samat kuin yleistetyssä Pareto-jakaumassa [18].

Käyttämästämme tanskalaisesta tulipalodatasta yleistetyllä Pareto-jakaumalla lasketut tulokset on esitelty taulukossa 4.1 alla ja laskentaan käytetty koodi on liitetiedoissa kuvassa A.3.

	5%	2.5%	1%
VaR	9.23	14.8	27.44
ES	28	44.4	81.8

**Taulukko 4.1.** VaR ja ES arvot tanskalaiselle tulipalodatalle (milj. DKK).

Taulukosta 4.1 nähdään esimerkiksi, että 9,23 miljoonan Tanskan kruunun suuruiset korvaukset aiheutuvat 5% todennäköisyydellä (VaR rivi) tai, että odotetun korvauksen kokonaissuuruus korkeintaan 1% todennäköisyydellä on 81,8 miljoonaa Tanskan kruunua (ES rivi). Tällä tavoin saamme siis laskettua konkreettisen riskin halutuilla todennäköisyyksillä hyödyntäen yleistetyllä Pareto-jakaumalla estimoituja parametrejä. Mikäli olettaisimme, että data olisi normaalisti jakautunutta, vastaavat arvot olisivat 1% todennäköisyydellä 23,2 miljoonaa Tanskan kruunua (VaR) ja vain 26,1 miljoonaa Tanskan kruunua (ES). Tämä ero kiteyttääkin hyvin työn alussa mainittua eroa siitä, miten esimerkiksi vakuutusyhtiöt voivat varautua mahdollisesti tuleviin, merkittäviin, vakuutuskorvauksiin.

## 5. YHTEENVETO

Työssä esiteltiin, määriteltiin ja todistettiin paksuhäntäinen jakauma olemassa olevaan kirjallisuuteen tukeutuen. Aluksi esiteltiin paksuhäntäisen jakauman käyttökohteita sekä muutamia sen jakaumia. Tämän jälkeen, teoriaosuudessa, paksuhäntäinen jakauma määriteltiin matemaattisesti. Määritelmät kuvasivat paksuhäntäisen jakauman ominaisuuksia erityisesti sen (oikean) hännän päässä. Teoriaosuus sisälsi myös määritelmiä kokoavat lauseet ja näiden matemaattiset todistukset. Teoriaosuuden lopussa määriteltiin vielä tarkemmin yksi paksuhäntäinen jakauma, yleistetty Pareto-jakauma, ja esiteltiin suppeammin myös kaksi muuta tunnettua paksuhäntäistä jakaumaa, Cauchy-jakauma sekä Burr-jakauma.

Mallinnusvaiheessa testattiin yleisen Pareto-jakauman sopivuutta aineistoon, joka sisältää dataa Tanskassa vuodesta 1980 vuoteen 1990 toteutuneista tulipaloista aiheutuneista palovakuutuskorvauksista. Mallinnusvaiheessa todettiin ensin käytetyn datan paksuhäntäisyys, hyödyntämällä keskimääräisen ylityksen (mean excess) funktiota. Tämän jälkeen tarkasteltiin, olemassa olevaan kirjallisuuteen tukeutuen, yleistetyn Pareto-jakauman parametrien estimointia suurimman uskottavuuden menetelmällä matemaattisesti. Itse mallinnus suoritettiin R-ohjelmistoa hyödyntäen. Jakaumalle saatiin onnistuneesti estimoitua sopivat parametrit. Lopputuloksena tunnistettiin, että yleistetyllä Pareto-jakaumalla voidaan kuvata erityisesti kyseisen aineiston niitä havaintoja, jotka ylittävät korvaussuuruudeltaan 2,4 miljoonaa Tanskan kruunua.

Työssä tarkastelu rajattiin kattamaan vain jakaumien (oikean) hännän ominaisuuksia, sillä paksuhäntäisiä jakaumia käytetäänkin juuri näiden häntäpäähän havaintojen arviointiin. Tämän takia jatkotutkimusaiheena voisikin olla hyvä pohtia paksuhäntäisten jakaumien soveltuvuutta koko aineiston tarkasteluun tai vaihtoehtoisesti hyödyntää aineiston tarkastelussa useampaa eri jakaumaa. Aineiston käsittelyn aikana havaittiin, ettei yleistetty Pareto-jakauma ole välttämättä paras mahdollinen jakauma kuvaamaan aineiston yleisimpien tapausten (pienen korvausummien) muodostamaa jakaumaa, joten voisikin olla syytä pohtia, mikä jakauma parhaiten tukisi yleistetyn Pareto-jakauman heikkouksia/rajoituksia.

Vastaava mallinnus voitaisiin lisäksi toteuttaa samalta alueelta ja samalla metodilla kerätyillä korvausummilla, mutta uudemmalla datalla. Mallinuksen jälkeen tuloksia voitaisiin verrata tämän työn tuloksiin ja pohtia mahdollisten erojen syitä. Onko esimerkiksi ilmastomuutos tai kasvaneet korjauskustannukset vaikuttaneet palojen esiintymistiheyteen tai maksettujen korvausten suuruuteen? Lisäksi voitaisiin tutkia kuinka nämä mahdolliset muutokset vaikuttavat yleistetyn Pareto-jakauman parametreihin vai ovatko tässä työssä estimoidut parametrit edelleen päteviä.

Kolmas ehdotus jatkotutkimukselle olisi liittää vastaavaanlainen ilmatoriskin paksuhäntäinen mallinnus vakuutusyhtiön hinnoitteluun. Kuinka vakuutusyhtiöt käytännössä siirtävät mallinnuksessa estimoidun ilmatoriskin tuotteidensa hinnoitteluun vai laskevatko he tietyn luottamusvälin, jonka ulkopuolisen riskin he ovat itse valmiita kantamaan? Lisäksi voisi selvittää kuinka yllä mainittu ongelma siitä, ettei paksuhäntäiset jakaumat ole parhaita jakauman alkupään havaintojen arviointiin, on ratkaistu.

## LÄHTEET

- [1] Ayman Alzaatreh et al. *The generalized Cauchy family of distributions with applications*. 2016. URL: <https://jsdajournal.springeropen.com/articles/10.1186/s40488-016-0050-3>.
- [2] Philippe Artzner et al. *COHERENT MEASURES OF RISK*. 1998. URL: <https://people.math.ethz.ch/~delbaen/ftp/preprints/CoherentMF.pdf>.
- [3] A.A Balkema ja L de Haan. *Residual Life Time at Great Age*. Version Vol.2 (5), 792-804. 1974. URL: <https://www-jstor-org.libproxy.tuni.fi/stable/2959306?sid=primo&seq=1>.
- [4] Maurice C. Bryson. *Heavy-Tailed Distributions: Properties and Tests*. Version Vol.16 (1), s.61-68. 1974. URL: <https://www-tandfonline-com.libproxy.tuni.fi/doi/epdf/10.1080/00401706.1974.10489150?needAccess=true>.
- [5] James Chen. *Conditional Value at Risk (CVar): Definition, Uses, Formula*. 2020. URL: [https://www.investopedia.com/terms/c/conditional\\_value\\_at\\_risk.asp](https://www.investopedia.com/terms/c/conditional_value_at_risk.asp).
- [6] Marc N. Conte ja David L. Kelly. *An imperfect storm: Fat-tailed tropical cyclone damages, insurance, and climate policy*. 2018. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0095069616302315#preview-section-introduction>.
- [7] Roger M. Cooke, Daan Nieboer ja Jolanta Misiewicz. *Fat-Tailed Distributions: Data, Diagnostics and Dependence*. Resources for the Future Department of Mathematics, Delft University of Technology. 2011. URL: [https://rogermcooke.net/rogermcooke\\_files/RFF-DP-11-19-REV-HeavyTails.pdf](https://rogermcooke.net/rogermcooke_files/RFF-DP-11-19-REV-HeavyTails.pdf).
- [8] R. Deidda. *A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series*. 2010. URL: <https://hess.copernicus.org/articles/14/2559/2010/>.
- [9] R Documentation. *R-datapaketin kuvaus*. URL: <https://search.r-project.org/CRAN/refmans/evir/html/danish.html>.
- [10] *e-Handbook of Statistical Methods*. National Institute of Standards and Technology. viitattu: 2023. URL: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda3663.htm>.
- [11] Paul Embrechts, Claudia Klüppelberg ja Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance*. 2013. URL: <https://ebookcentral.proquest.com/lib/tampere/reader.action?docID=3088224&ppg=1#>.
- [12] Sergey Foss, Dimitry Korshunov ja Stan Zachary. *An Introduction to Heavy-Tailed and Subexponential Distributions*. 2011. URL: <https://link.springer-com.libproxy.tuni.fi/book/10.1007/978-1-4419-9473-8>.

- [13] Yari Gholamhossein ja Tondpour Zahra. *The new Burr distribution and its application*. 2017. URL: <https://link.springer.com/article/10.1007/s40096-016-0203-z#Sec7>.
- [14] Souvik Ghosh ja Sidney Resnick. *A discussion on mean excess plots*. 2010. URL: <https://www.sciencedirect.com/science/article/pii/S0304414910001079#fig2>.
- [15] A R Hakim, I Fithriani ja M Novita. *Properties of Burr distribution and its application to heavy-tailed survival time data*. 2021. URL: <https://iopscience.iop.org/article/10.1088/1742-6596/1725/1/012016/pdf>.
- [16] David R. Harper. *What Is Value at Risk (VaR) and How to Calculate It?* 2023. URL: <https://www.investopedia.com/articles/04/092904.asp>.
- [17] Devandra Kumar. *The Burr Type Xii Distribution with Some Statistical Properties*. 2017. URL: [https://www.researchgate.net/publication/316622805\\_The\\_Burr\\_Type\\_Xii\\_Distribution\\_with\\_Some\\_Statistical\\_Properties](https://www.researchgate.net/publication/316622805_The_Burr_Type_Xii_Distribution_with_Some_Statistical_Properties).
- [18] Wo-Chiang Lee. *Fitting the generalized Pareto distribution to commercial fire loss severity: evidence from Taiwan*. 2012. URL: <https://core.ac.uk/download/pdf/225216209.pdf>.
- [19] B Li, Jianguo Lin ja Yao Xiuming. *Characteristics of the optimization of creep constitutive equations*. 2003. URL: [https://www.researchgate.net/publication/281580270\\_Characteristics\\_of\\_the\\_optimization\\_of\\_creep\\_constitutive\\_equations](https://www.researchgate.net/publication/281580270_Characteristics_of_the_optimization_of_creep_constitutive_equations).
- [20] Mikko Mattila, Markus Kaakinen ja Noora Ellonen. *Regressioanalyysi*. 2003. URL: <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/regressio/arviointi/>.
- [21] James Millington, George L.W. Perry ja Bruce D. Malamud. *Models, data mechanisms: Quantifying wildfire regimes*. 2006. URL: [https://www.researchgate.net/publication/224011819\\_Models\\_data\\_mechanisms\\_Quantifying\\_wildfire\\_regimes](https://www.researchgate.net/publication/224011819_Models_data_mechanisms_Quantifying_wildfire_regimes).
- [22] Jayakrishnan Nair, Adam Wierman ja Zwart Bert. *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. 2022. URL: <https://doi.org/10.1017/9781009053730>.
- [23] Minh H. Pham, Chris Tsokos ja Bong-Jin Choi. *Maximum Likelihood Estimation for the Generalized Pareto Distribution and Goodness-of-Fit Test with Censored Data*. Version Vol.17 (2). 2019. URL: <https://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=2608&context=jmasm>.
- [24] James Pickands. *Statistical Inference Using Extreme Order Statistics*. Version vol.3 (1), p.119-131. 1975. URL: <https://www.jstor-org.libproxy.tuni.fi/stable/2958083?sid=primo>.
- [25] Analyst Prep. *Extensions of VaR*. 2021. URL: <https://analystprep.com/study-notes/cfa-level-2/describe-extensions-of-var/>.
- [26] William J. Reed ja Kevin S. McKelvey. *Power-law behaviour and parametric models for the size-distribution of forest fires*. 2002. URL: <https://www-sciencedirect-com.libproxy.tuni.fi/science/article/pii/S0304380001004835>.

- [27] Nur Hanim Mohd Salleh ja Husna Hasan. *GENERALIZED PARETO DISTRIBUTION FOR EXTREME TEMPERATURES IN PENINSULAR MALAYSIA*. 2018. URL: <http://www.sci-int.com/pdf/636531658004820921.pdf>.
- [28] Richard L. Smith. *Estimating Tails of Probability Distributions*. Version Vol.15 (3), s.1174-1207. 1987. URL: <https://www-jstor-org.libproxy.tuni.fi/stable/2241822?sid=primo&seq=2>.
- [29] David Strauss, Larry Bednar ja Romain Mees. *Do One Percent of Forest Fires Cause Ninety-Nine Percent of the Damage*. 1989. URL: [https://www.fs.usda.gov/psw/publications/strauss/psw\\_1989\\_strauss001.pdf](https://www.fs.usda.gov/psw/publications/strauss/psw_1989_strauss001.pdf).

## LIITE A: KOODI

```
1 library(evir)
2 library(fitdistrplus)
3
4 data(danish)
5 X <- danish
6
7 xi <- c()
8 beta <- c()
9 threshold <- c()
10 nextrem <- c()
11
12 keep <- 0
13
14 # arvioidaan jokaisella u:lle GDP:n parametrit
15
16 for (i in 1:ceiling((1-keep)*length(X))) {
17   skip_to_next <- FALSE
18   tryCatch(pareto <- gpd(X, nextremes = i, method = "ml"),
19     error = function(e) {skip_to_next <-< TRUE})
20   if(skip_to_next) {
21     next
22   }
23   threshold[i] <- pareto$threshold
24   xi[i] <- pareto$par.ests[1]
25   beta[i] <- pareto$par.ests[2]
26   nextrem[i] <- pareto$n.exceed
27 }
28
29 xi <- na.omit(xi)
30 beta <- na.omit(beta)
31 threshold <- na.omit(threshold)
32 nextrem <- na.omit(nextrem)
```

**Kuva A.1.** Parametrien estimointi ja datansovitus osa1.

```

33
34 # Arvioitu GDP
35
36 val <- ceiling(length(X)/10)
37 varri <- c()
38 indeksi <- c()
39 for (i in 1:(length(xi)-val +1)) {
40   z <- i
41   y <- i + val -1
42   varri[i] <- var(xi[z:y])
43   indeksi[i] <- z
44 }
45
46 # Lasketaan paras threshold (raja-arvo)
47 alaraja <- threshold[indeksi[order(varri)[1]]]
48 ylaraja <- threshold[indeksi[order(varri)[1]] + val -1]
49 u <- 0.5 * (alaraja + ylaraja)
50
51 pareto <- gpd(X, threshold = u, method = "ml")
52   #metodi "ml" tarkoittaa suurimman uskottavuuden menetelmää
53 plot(pareto)
54
55 # Make a plot selection (or 0 to exit):
56
57 #1: plot: Excess Distribution #
58 #2: plot: Tail of Underlying Distribution
59 #3: plot: Scatterplot of Residuals
60 #4: plot: QQplot of Residuals

```

**Kuva A.2.** Parametrien estimointi ja datansovitus osa2.

```

61
62
63 aThreshold <- pareto$threshold
64 aXi <- pareto$par.ests[1]
65 aBeta <- pareto$par.ests[2]
66 aNextrem <- pareto$n.exceed
67
68 p <- c(0.95, 0.975, 0.99)
69
70 VaR <- aThreshold + (aBeta/aXi)*((((length(X)/aNextrem)*(1-p))^(1-aXi))-1)
71 ES <- VaR * ((1/(1-aXi))+((aBeta-aXi*aThreshold)/((1-aXi)*VaR)))
72
73 out <- data.frame(VaR, ES, p) #VaR ja ES tulokset

```

**Kuva A.3.** VaR ja ES laskentakoodi.