



Ääriarvoteoria ja Value-at-Risk ja niiden soveltaminen  
riskienhallintaan

Jelena Vuorinen

Pro gradu -tutkielma  
Toukokuu 2020

MATEMATIIKAN JA TILASTOTIETEEN LAITOS  
TURUN YLIOPISTO

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä

TURUN YLIOPISTO

Matematiikan ja tilastotieteen laitos

VUORINEN, JELENA: Ääriarvoteoria ja Value-at-Risk ja niiden soveltamisen riskienhallintaan

Pro gradu -tutkielma, 47 s., 12 liites.

Sovellettu matematiikka

Toukokuu 2020

---

Riskin arvioimiselle ja hallitsemiselle on tarpeita monilla eri aloilla. Klassisia esimerkkejä ovat esimerkiksi sijoitusportfolion tappion tai tulvaveden korkeuden arvioiminen. Mitä tarkemmin äärimmäisten havaintojen esiintymistodennäköisyyksiä voidaan arvioida, sitä helpompi niihin on yrittää varautua.

Value-at-Risk on alunperin markkinariskin mittaamiseen kehitetty menetelmä. Sitä käytetään myös esimerkiksi vakuutusyhtiöiden vakavaraisuuden sääntelyyn. Ääriarvoteoria on yksi käytetyimmistä lähestymistavoista riskin arvioimiseen. Ääriarvoteoriaa ja Value-at-Riskia yhdistämällä voidaan tavanomaisia tilastollisia menetelmiä tarkemmin arvioida todennäköisyyttä sille, että äärimmäiseksi luokiteltava havainto toteutuu.

Tutkielman tavoitteena on esitellä Value-at-Riskia ja ääriarvoteoriaa sekä sitä, miten niitä voidaan soveltaa käytäntöön. Yhtenä käytännön sovelluskohteena esitellään lyhyesti Solvenssi II -raportointia. Ääriarvoteorian soveltamiseen käytetään blokkimaksimimenetelmää ja ylitemenetelmää. Työssä esiteltyjen menetelmien avulla mallinnetaan äärimmäisten havaintojen todennäköisyysjakaumia  $R$ -ohjelmointikielellä. Aineistona käytetään Tilastokeskuksen julkaisemaa tilastoa kuolemansyistä. Tilastosta käytetään kuolleiden kokonaismäärää. Koska käytetyssä aineistossa on niin vähän havaintoja, simuloidaan vuosittaiset kuolemat kuukausitasolle mallinnuksen helpottamiseksi.

Sekä blokkimaksimi- että ylitemenetelmää on melko helppo soveltaa  $R$ -ohjelmointikielellä ja ne antavat keskenään hieman erilaisia tuloksia. Ylitemenetelmää käytettäessä myös kynnsarvon valinta vaikuttaa tuloksiin ja siksi sen valintaan on tärkeää kiinnittää erityistä huomiota. Sopivan arvon löytäminen voi olla hankalaa, ja siksi on hyvä valita muutamia kynnsarvoja ja vertailla niiden antamia malleja keskenään.

Koska pohjana käytetyn aineiston havaintovälille ei osu juurikaan poikkeavia vuosia kuolleiden määrän suhteen, voidaan siitä simuloitua aineistoa kuvaavien mallien ajatella aliarvioivan riskiä. Mitä pidemmälle aineiston ulkopuolelle ekstrapoloidaan, sitä enemmän epävarmuutta mallit aina sisältävät.



# Sisältö

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Value-at-Risk</b>	<b>2</b>
2.1	Historia . . . . .	2
2.2	Määritelmiä . . . . .	2
2.3	Value-at-Risk -menetelmät . . . . .	3
2.3.1	Parametrinen Value-at-Risk . . . . .	3
2.3.2	Historiallinen simulointi . . . . .	4
2.3.3	Monte Carlo -simulointi . . . . .	5
2.4	Solvenssi II . . . . .	5
<b>3</b>	<b>Ääriarvoteoria</b>	<b>7</b>
3.1	Johdanto ääriarvoteoriaan . . . . .	7
3.2	Aputuloksia satunnaismuuttujajonojen maksimeista . . . . .	8
3.3	Ääriarvomenetelmät . . . . .	9
3.3.1	Ääriarvojakaumat ja blokkimaksimimenetelmä . . . . .	9
3.3.2	Yleistetty Pareto-jakauma ja ylitemenetelmä . . . . .	14
<b>4</b>	<b>Ääriarvoteoria yhdistettynä Value-at-Risktiin</b>	<b>17</b>
4.1	Toistumisjakso ja toistumistaso . . . . .	17
4.1.1	Yleistetty ääriarvojakauma . . . . .	18
4.1.2	Yleistetty Pareto-jakauma . . . . .	19
4.2	Mallinnus . . . . .	22
4.2.1	Aineiston esittely . . . . .	22
4.2.2	Blokkimaksimimenetelmä . . . . .	23
4.2.3	Ylitemenetelmä . . . . .	26
4.2.4	Mallien vertailu ja yhteenveto . . . . .	39
4.3	Johtopäätökset . . . . .	44
	<b>Lähteet</b>	<b>46</b>
<b>A</b>	<b>Käytetty R-ohjelmointikielen koodi</b>	<b>48</b>

# 1 Johdanto

Riskin arvioiminen on tärkeää monella alalla. Esimerkiksi vakuutusala on velvoittanut Solvenssi II, joka on Euroopan parlamentin ja neuvoston direktiivi vakuutus- ja jälleenvakuutustoiminnan aloittamisesta ja harjoittamisesta. Osana Solvenssi II -raportointia vakuutusyhtiöiden tulee arvioida vakavaraisuuteensa kohdistuvia riskejä.

Value-at-Risk on 1900-luvulla alun perin markkinariskin mittaamiseen kehitetty menetelmä, jota nykyisin käytetään myös muun muassa edellä mainitussa Solvenssi II -raportoinnissa. Value-at-Riskin mittaamiseksi täytyy pystyä arvioimaan harvinaisten tapahtumien todennäköisyyksiä. Koska havainnot ovat äärimmäisiä, niitä on vaikea ennustaa. Perinteiset tilastolliset estimointimenetelmät mallintavat tarkasteltavan ilmiön yleisiä havaintoja. Todennäköisyysjakaumien hännät, joissa äärimmäiset havainnot ovat, jäävät perinteisillä estimointimenetelmillä epätarkemmiksi. Tällöin Value-at-Riskin estimaateistakin tulee epätarkkoja.

Ääriarvoteoria on 1900-luvulla kehitetty tilastollinen menetelmä, joka mahdollistaa äärimmäisten tapahtumien ennustamisen tarkemmin. Ääriarvoteorian mukaan todennäköisyysjakaumien häntiä mallinetaan erikseen muusta jakaumasta. Keskittymällä vain äärimmäisiin havaintoihin saadaan tavallista tarkempi arvio niiden ilmentymistodennäköisyyksistä. Ääriarvoteoria kehitettiin alun perin tulvavedenkorkeuksien tutkimiseen. Nykyään sillä kuten Value-at-Riskilläkin on sovelluksia useilla aloilla. Ääriarvoteorian soveltamiseen käytetään yleensä joko blokkimaksimimenetelmää tai ylitemenetelmää.

Tässä työssä esitellään Value-at-Riskin teoriaa ja ääriarvoteoriaa. Työssä esitellään myös hieman Solvenssi II -raportointia, sillä se on yksi Value-at-Riskin konkreettisista käyttökohteista. Lopuksi tehdään oma mallinnus esiteltyjen menetelmien soveltamiseksi käytäntöön. Aineistona käytetään Tilastokeskuksen tilastoa kuolemien määrästä vuosittain. Aineisto simuloidaan kuukausitasolle, jotta saadaan enemmän havaintoja, joiden pohjalta ääriarvoteorian malleja voidaan rakentaa.

## 2 Value-at-Risk

### 2.1 Historia

Value-at-Risk, josta käytetään lyhennettä VaR, on markkinariskin mittaamiseen kehitetty menetelmä. Variaatiota siitä on käytetty New Yorkin pörssissä jo 1920-luvulla, vaikka virallisesti termi Value-at-Risk on otettu käyttöön vasta 1990-luvun alussa. Ensimmäisen kerran termiä käytettiin vuonna 1993 julkaistussa teoksessa *Derivatives: Practices and Principles*, jonka on julkaissut rahoitusalan johtavien instituutioiden johtajista ja akateemikoista koostunut ryhmä nimeltään Group of 30 [7]. VaR oli lähinnä finanssiyritysten riskianalyttikoiden työkalu, mutta J.P.Morganin vuonna 1994 julkaiseman ilmaisen RiskMetrics-palvelun [11] myötä sen käyttö yleistyi.

### 2.2 Määritelmiä

Value-at-Risk on alunperin määritelty markkinariskin mittaamiseen, mutta tässä työssä sitä käytetään mittaamaan riskiä yleisemmin. Esitellään Choudhryn (2013) [10] käyttämä määritelmä Value-at-Riskille.

**Maaritelma 2.1** (VaR). VaR on suurin todennäköinen vahinko, joka tapahtuu määritellyllä aikavälillä ja varmuusasteella (degree of certainty).

Value-at-Riskin voidaan ajatella koostuvan kolmesta rakenneosasta. Ensimmäinen niistä on tarkastelu-aika, eli aikaväli, jolla tapahtuvaa suurinta todennäköistä vahinkoa arvioidaan. Finanssialalla tarkastelu-aika on yleensä yksi päivä, kun taas vakuutuslalla tarkastellaan tavallisesti pidempiä aikavälejä. Toinen rakenneosa on varmuusväli, eli se luottamustaso, joka Value-at-Risk -estimaatille asetetaan. Tyypillisesti käytetään joko 95% tai 99% luottamustasoa. Jotta luottamustaso voidaan laskea, pitää kolmantena osana määrittellä todennäköisyysjakauma, jota tarkasteltavan ilmiön oletetaan noudattavan.

Muodollisesti VaR voidaan määrittellä seuraavasti kuten Leppisaari (2013) [9] on tehnyt. Olkoon  $0 < \alpha < 1$ . VaR luottamustasolla  $\alpha$  on pienin reaaliluku  $l$ , jolle todennäköisyys, että havaittu tappio  $L$  ylittää arvon  $l$  on pienempi

tai yhtä suuri kuin  $1 - \alpha$ . Siis

$$\text{VaR}_\alpha = \inf \{l \in \mathbb{R} : \mathbb{P}(L < l) \leq 1 - \alpha\} = \inf \{l \in \mathbb{R} : F_L \geq \alpha\} = F_L^{\leftarrow}(\alpha),$$

missä merkintä  $F^{\leftarrow}$  tarkoittaa funktion  $F$  yleistettyä käänteisfunktioita. Eli  $F_L^{\leftarrow}(\alpha)$  on funktion  $F_L$   $\alpha$ -kvantiili. Jatkuvalle jakaumalle  $F^{\leftarrow} = F^{-1}$ .

## 2.3 Value-at-Risk -menetelmät

Value-at-Risk voidaan laskea usealla eri menetelmällä, jolloin myös sen saama arvo saattaa vaihdella menetelmästä riippuen. Seuraavaksi esitellään kolme yleisimmin käytössä olevaa menetelmää VaR:in laskemiseksi. Ne ovat *parametrinen Value-at-Risk*, *historiallinen simulointi* ja *Monte Carlo -simulointi*.

### 2.3.1 Parametrinen Value-at-Risk

Parametrinen Value-at-Riskin esittely pohjautuu Dowdin ja Blaken (2006) [3] artikkeliin. Parametrinen VaR perustuu oletukseen, että tarkasteltavalla ilmiöllä on todennäköisyysjakauma joillain parametreilla. Kun todennäköisyysjakauma ja sen parametrit ovat selvillä, saadaan VaR todennäköisyysjakauman kvantileista. Menetelmä on periaatteessa yksinkertainen, mutta siinä on omat ongelmansa. Todennäköisyysfunktioita selvittäessä täytyy ottaa huomioon, onko ilmiö riippuvainen historiastaan tai muista satunnaisista ilmiöistä. Esimerkiksi omaisuusvahinkojen todennäköisyys on riippuvainen luonnonilmiöistä, sillä myrskyt ja tulvat ovat näiden vahinkojen yleisiä aiheuttajia. Tällöin aineistolle pitää valita ehdollinen jakauma. Aiemman tutkimuksen perusteella joillekin ilmiöille sopivat todennäköisyysjakaumat ovat jo tiedossa, jolloin tehtäväksi jää vain parametrien selvittäminen. Tämä tapahtuu yleensä estimoimalla niitä olemassaolevasta aineistosta. Usein käytettyjä parametrien estimointimenetelmiä ovat *momenttimenetelmä*, *suurimman uskottavuuden menetelmä* ja *pienimmän neliösumman menetelmä*. Kun ilmiötä kuvaava todennäköisyysfunktio parametreineen on selvitetty, antaa haluttua luottamustasoa vastaava kvantiili VaR:in arvon. Todennäköisyysfunktioita ei yleensä tiedetä tarkasti, joten menetelmää käytettäessä pitää muistaa että



parametrien estimaatit ja VaR eivät ole eksakteja. Niille voidaan kuitenkin laskea luottamusvälit.

Dowd ja Blake (2006) [3] mukaan parametrisen VaRin heikkous on, että sen soveltaminen vaatii yleensä huomattavia oletuksia ja yksinkertaistuksia. Se toimii käytännössä vain hyvin yksinkertaisille tapauksille ja muulloin se on erittäin epäluotettava. Väärän todennäköisyysjakauman valinta voi joutaa vakaviin virheisiin riskin arvioinnissa. Estimoitavan todennäköisyysfunktion valinta voi suppealla tai toisistaan riippuvista havainnoista koostuvalla aineistolla olla erittäin hankalaa.

### 2.3.2 Historiallinen simulointi

Historiallinen simulointi on parametrittomista VaR:in määrittelymenetelmistä yleisin. Siinä ennustetaan tulevaisuutta menneiden havaintojen perusteella. Historiallisen simuloinnin käyttöön perusmuodossaan liittyy oletus, että tarkasteltavalle ilmiölle tulevaisuuden havainnot vastaavat menneisyyttä hyvin. Menetelmässä historiallisesta aineistosta havaitut frekvenssit kullekin muuttujan arvolle toimivat estimaattina tuleville havainnoille samasta ilmiöstä. Tarkastellaan esimerkiksi keskilämpötilan eroa kahden peräkkäisen päivän välillä. Luodaan ilmiön todennäköisyysjakaumaa simuloiva histogrammi siten, että menneisyyden havainnot jaetaan havaintoväleille jotka on laitettu suuruusjärjestykseen. Kuhunkin väliin osuvien havaintojen kappalemäärä kertoo kyseisen välin yleisyydestä. Lopuksi vain luetaan haluttua luottamustasoa vastaava kvantiili histogrammista VaR:in estimaatiksi. Historiallinen simulointi on erittäin yksinkertainen toteuttaa.

Holtonin (2014) [8] mukaan menetelmän heikkoutena on, että laadukasta aineistoa täytyy olla kerättynä tarpeeksi, jotta menetelmällä saatu VaR:in estimaatti olisi mahdollisimman virheetön. Myös sillä on estimaatin hyvyyden kannalta merkitystä, miltä aikaväliltä aineistoa käytetään. Esimerkiksi laskettaessa VaR:ia historiallisella simuloinnilla älylaitteisiin kohdistuneiden korvausten määrälle ei kannata käyttää kovin vanhaa aineistoa. Ennen älylaitteita oli käytössä paljon vähemmän, joten esimerkiksi kymmenen vuoden takaiset vahinkomäärät eivät ennusta tulevaisuuden vahinkoja kovin hyvin.

Historiallisen simuloinnin ongelmana on, että yleensä ilmiölle äärimmäisiä havaintoja on aineistossa vain vähän jolloin VaR:in estimaatista tulee epätarkka. Äärimmäisten havaintojen puute voi tehdä VaR:ista liian optimistisen, tai vastaavasti yksi erittäin poikkeuksellinen havainto voi tehdä estimaatista aivan liian pessimistisen. Kun halutaan tarkastella nimenomaan äärimmäisiä havaintoja, ovat epäparametriset menetelmät usein riittämättömiä.

### 2.3.3 Monte Carlo -simulointi

Monte Carlo -simulointia käytetään yleisesti stokastisten ilmiöiden mallintamisessa. Menetelmää on esitellyt muun muassa Owen A.B. (2013) [12]. Menetelmässä simuloidaan tarkasteltavaa ilmiötä kuvaavaa aineistoa satunnaislukujen avulla. Yleinen tapa on mallintaa ilmiöön vaikuttavia tekijöitä sen sijaan että mallinnettaisiin ilmiötä suoraan. Ensin selvitetään kaikki ilmiöön vaikuttavat tekijät ja niiden mahdolliset suhteet toisiinsa. Tämän jälkeen vaikuttavista tekijöistä määriteltyjen muuttujien todennäköisyysjakaumat selvitetään. Kun jokaiselle muuttujalle generoidaan satunnaislukujen avulla toteutuma, saadaan toteutumat yhdistämällä luotua simuloitu näyte tutkittavasta ilmiöstä. Kun kaikki simuloidut näytteet kootaan yhteen, antaa niiden jakauma approksimaation tarkasteltavalle ilmiölle. Tällöin VaR saadaan määriteltyä kuten historiallisen simuloinninkin tapauksessa vaitsemalla histogrammista haluttu kvantiili.

Menetelmän vahvuutena on, että tarkasteltavan ilmiön todellista jakaumaa ei tarvitse tuntea. Toisaalta ilmiöön vaikuttavien tekijöiden tunnistaminen ja niiden jakaumien tuottaminen voi olla hankalaa. Hyvään tarkkuuteen pääsemiseen tarvitaan kuitenkin suuri otos simuloituja näytteitä, joista jokaista varten luodaan useita satunnaislukuja. Tällöin monimutkaisten ilmiöiden simuloiminen saattaa käyttää niin paljon laskentatehoa, ettei menetelmän käyttö ole resurssien puutteessa mahdollista.

## 2.4 Solvenssi II

Solvenssi II on Finanssivalvonnan [5] mukaan Euroopan parlamentin ja neuvoston direktiivi vakuutus- ja jälleenvakuutustoiminnan aloittamisesta

ja harjoittamisesta. Se on annettu marraskuussa 2009 ja otettu käyttöön 1.1.2016. Sen valvojana toimii Euroopan vakuutus- ja lisäeläkevalvontaviranomainen EIOPA (The European Insurance and Occupational Pensions Authority), joka raportoi Euroopan parlamentille ja komissiolle.

Vakavaraisuusjärjestelmä Solvenssi II on korvannut aiemman Solvenssi I:n, joka on peräisin 1970-luvulta. Finanssivalvonnan mukaan Solvenssi II:n on tarkoitus edistää yritysten omaa riskienhallintaa ja pääomien tehokasta hyödyntämistä sekä edistää EU:n sisäistä kilpailua. Se sisältää ohjeita vakuutus- ja jälleenvakuutusyhtiöille niiden toiminnan ja sen valvonnan yhtenäistämiseksi.

Solvenssi II:n ohjeiden voidaan sanoa muodostavan kolme pilaria. Ensimmäinen sisältää kvantitatiiviset vaatimukset ja pitää sisällään esimerkiksi yhtiöiden vähimmäispääomavaatimuksen (Minimal Capital Requirement, MCR) sekä vakavaraisuuspääomavaatimuksen (Solvency Capital Requirement, SCR). Toinen pilari muodostuu kvantitatiivisista vaatimuksista, jotka koskevat vakuutusyhtiöiden hallintojärjestelmiä. Viimeinen pilari sisältää ohjeita ja vaatimuksia vakuutusyhtiöiden valvontaraportointiin ja tietojen julkistamiseen ja näin ollen niiden toiminnan läpinäkyvyyteen liittyen.

Solvenssi II siis muun muassa sääntelee vakuutusyhtiöiden vakavaraisuutta asettamalla vaatimuksia niiden pääomalle. Tiaisen (2011) [6] ja Euroopan komission [4] mukaan näiden vaatimustasojen laskemiseen käytetään Value-at-Riskiä. Vakavaraisuussäätely perustuu vakavaraisuuspääomavaatimukseen (SCR) sekä vähimmäispääomavaatimukseen (MCR). Ne toimivat eräänlaisina hälytystasoina vakuutus- ja jälleenvakuutusyrityksille, ja tasojen välillä on eriasteisia viranomaistoimia, joilla puututaan yrityksen toimintaan ja yritetään korjata sitä. Mikäli yrityksen oma perusvarallisuus jää viranomaisten väliintulosta huolimatta alle vähimmäispääomavaatimuksen, eikä yritys pysty antamaan realistista lyhyen aikavälin rahoitussuunnitelmaa sen korjaamiseksi, sen toimilupa perutaan. Nämä pääomavaatimukset on määriteltä suojamaan vakuutuksenottajia.

Korkeampi pääomavaatimuksista on siis vakavaraisuuspääomavaatimus SCR, joka ottaa huomioon kaikki vakuutusyhtiön vakavaraisuutta uhkaavat riskit ja niiden yhteisvaikutukset. Se perustuu Value-at-Riskiin, ja siinä vaa-

ditaan 99.5%:n luottamustaso yhden vuoden tarkasteluaajalla. SCR:n määrittämiseksi on vakuutusyhtiöiden käyttöön kehitetty yleinen malli, jonka on tarkoitus kuvata useimpien vakuutusyhtiöiden riskiprofilia. Jos yrityksen riskiprofilin tiedetään olevan yleisestä mallista poikkeava, SCR voidaan myös laskea kokonaan tai osittain yrityksen sisäisellä mallilla. Finanssivalvonnan tulee kuitenkin hyväksyä tämä sisäinen malli.

Alempi pääomavaatimuksista on vähimmäispääomavaatimus MCR. Sen laskemiseen käytetään lineaarista funktiota, joka kalibroidaan yrityksen omien perusvarojen Value-at-Riskin arvoa vastaavalle tasolle. MCR:lle Value-at-Risk lasketaan 85% luottamustasolla yhden vuoden ajanjaksolle. Lisäksi MCR:lle on asetettu absoluuttisia vähimmäistasoja, joiden suuruus riippuu vakuutusyrityksen tyypistä.

## 3 Ääriarvoteoria

### 3.1 Johdanto ääriarvoteoriaan

Kun halutaan arvioida riskiä, on ääriarvoteoria yksi tavallisimmista lähestymistavoista. Se on tilastollinen menetelmä, joka käsittelee todennäköisyysjakauman häntiä. Sen avulla on mahdollista ennustaa harvinaisia, äärimmäisiä tapahtumia ekstrapoloiden olemassaolevaa dataa. Alunperin ääriarvojen tilastollista analysointia käytettiin tulvavedenkorkeuksien tutkimiseksi, ja sille on biologisten ilmiöiden ennustamisen lisäksi monia sovelluskohteita esimerkiksi vakuutus- ja finanssialalla.

Ääriarvoteoriassa on keskeistä raja-arvolauseetta vastaava tulos, joka määrittelee otosmaksimien rajajakaumaa. Keskeinen raja-arvolause määrittelee toisistaan riippumattomien satunnaismuuttujien keskiarvon olevan likipitäen normaalijakautunut riippumatta sen alkioden alkuperäisestä jakaumasta. Ääriarvoteorian keskeinen lause puolestaan määrittelee satunnaismuuttujien maksimien rajajakaumien kuuluvan yhteen kolmesta jakaumaperheestä, joita kutsutaan *ääriarvojakaumiksi*. Tässä luvussa esitellään ääriarvoteoriaa ja tapoja sen soveltamiseen.

### 3.2 Aputuloksia satunnaismuuttujajonojen maksimeista

Leppisaari (2013) [9] on koonnut taustateoriaa satunnaismuuttujajonojen maksimeista johdatuksena ääriarvoteoriaan. Mukaillaan hänen käyttämäänsä esitystapaa. Tarkastellaan otosta  $X_1, \dots, X_n$ , jossa on  $n$  kappaletta riippumattomia ja samoin jakautuneita (iid, independent and identically distributed) satunnaismuuttujia, joilla on yhteinen kertymäfunktio (cdf, cumulative distribution function)  $F$ . Merkitään *otosmaksimia*  $M_n = \max(X_1, \dots, X_n)$ . Äärimmäisten ilmiöiden tarkastelua varten ollaan kiinnostuneita otosmaksimien jakaumasta. Sille on periaatteessa johdettavissa täsmällinen jakauma kaikille positiivisille kokonaisluvuille  $n$ . Täsmällinen jakauma pohjautuu otoksen kertymäfunktioon  $F$ , joka ei ole kuitenkaan usein tiedossa. Sen määritelmä on seuraava:

$$\begin{aligned} \mathbb{P}(M_n \leq x) &= \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \leq x) := F^n(x), \quad x \in \mathbb{R}, n \in \mathbb{N}. \end{aligned} \tag{1}$$

Tarkan määritelmän sijaan on mahdollista käyttää kertymäfunktiolle  $F$  estimaattia  $\hat{F}$ , joka saadaan määriteltyä havaitun aineiston pohjalta tilastollisilla menetelmillä. Käytännössä kertymäfunktioita on kuitenkin hankala estimoida, kun ollaan kiinnostuneita tähänastisia havaintoja suuremmista äärimmäisistä havainnoista. Pelkkien äärimmäisten havaintojen käyttäminen mallinnuksessa kuvaa paremmin kertymäfunktion hännän käyttäytymistä. Siksi onkin parempi mallintaa suoraan maksimien jakaumaa  $F^n$ .

Otosmaksimien havaitsemiseen on vain pieni todennäköisyys, eli ne sijoittuvat jakauman oikeaan häntään. Tarkastellaan seuraavaksi suureen  $M_n$  käyttäytymistä, kun  $n \rightarrow \infty$ . Olkoon  $F$  kertymäfunktio otokselle riippumattomia ja samoin jakautuneita satunnaismuuttujia ja  $x_F$  sen oikea päätepiste. Siis  $x_F = \sup\{x \in \mathbb{R} | F(x) \leq 1\}$ . Kun otoskoko kasvaa rajatta, saadaan

$$\mathbb{P}(M_n \leq x) = F^n(x) \xrightarrow{n \rightarrow \infty} 0, \quad x \leq x_F \leq \infty,$$

$$\mathbb{P}(M_n \leq x) = F^n(x) = 1, \quad x \geq x_F, x_F \leq \infty.$$

Otosmaksimin jakauma siis suppenee kertymäfunktioon, joka on porraskunktio. Näin saadaan  $M_n \xrightarrow{\mathbb{P}} x_F$ , eli otosmaksimi  $M_n$  suppenee stokastisesti pisteeseen  $x_F$ , kun  $n \rightarrow \infty$ . Koska jono  $M_n$  on ei-vähenevä parametrin  $n$  suhteen, myös  $M_n \xrightarrow{m.v.} x_F$  eli otosmaksimi konvergoi melkein varmasti todennäköisyysjakauman oikeaan pisteeseen  $x_F$ . Nyt tiedämme otosmaksimien jakaumasta  $F^n$ , että sen todennäköisyysmassa keskittyy pisteeseen  $x_F$ , kun  $n \rightarrow \infty$ . Tästä ei kuitenkaan ole vielä apua äärimmäisten arvojen esiintymistodennäköisyyksien mallintamisessa. Tarkastellaan lisäksi keskitettyjen ja normalisoitujen maksimien suppenemista jakauman suhteen, eli niiden heikkoa suppenemista.

Yleisellä tasolla maksimien heikkoa suppenemista ei voida tarkastella suoraan. Ei-degeneroitunutta rajajakaumaa ei välttämättä ole olemassa. Olkoon  $(a_n)$  ja  $(b_n)$  jonoja vakioita, joille  $a_n > 0 \forall n$ ,  $b_n \in \mathbb{R} \forall n$  ja  $n \geq 1$ . Nyt otosmaksimille  $M_n$  saadaan lineaarinen muunnos  $\frac{M_n - b_n}{a_n}$ . Tällöin joillain sopivilla vakiojonoilla  $a_n$  ja  $b_n$  löydetään kuitenkin varsin yleisten ehtojen ollessa voimassa rajajakauma  $G(x)$ . Eli

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

missä  $G$  on ei-degeneroitunut kertymäfunktio. Itseasiassa ääriarvoteorian keskeinen lause otosmaksimeille määrää, että jos rajajakauma  $G(x)$  on olemassa, sen täytyy kuulua yhteen kolmesta ääriarvojakaumaperheestä. Ääriarvojakaumista saadaan parametrisoimalla *yleistetty ääriarvojakauma*.

### 3.3 Ääriarvomenetelmät

#### 3.3.1 Ääriarvojakaumat ja blokkimaksimimenetelmä

Esitellään seuraavaksi ääriarvoteorian keskeisiä lauseita otosmaksimeille, sekä esitellään yleistetty ääriarvojakauma ja kolme jakaumaperhettä, joihin se voidaan jakaa. Lauseet ovat saaneet alkunsa R.A.Fisher ja L.H.C.Tippettin julkaisusta vuodelta 1928, ja niitä on täydentänyt B.V.Gnedenko vuonna

1943. Lauseiden todistukset sivuutetaan, sillä ne ovat melko pitkiä ja ne löytyvät de Haanin ja Ferreiran (2006) teoksesta [2].

**Lause 3.1** (Fisher-Tipett-Gnedenko, yleistetty ääriarvojakauma). Olkoon  $(X_i), i = 1, \dots, n$ , jono riippumattomia ja samoin jakautuneita satunnaismuuttujia. Jos on olemassa vakiojonot  $(a_n)$  ja  $(b_n)$ , joille  $a_n > 0 \forall n, b_n \in \mathbb{R} \forall n$  ja  $n \geq 1$ , ja on olemassa ei-degeneroitunut jakauma  $G_\gamma$  siten, että

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{M_n - b_n}{a_n} \leq x \right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x), x \in \mathbb{R}, \quad (2)$$

niin ääriarvojakaumien joukko on  $G_\gamma(ax + b)$ , jossa  $a > 0$  ja  $b \in \mathbb{R}$  ja

$$G_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & 1 + \gamma x > 0, \\ \exp(-e^{-x}), & \gamma = 0 \end{cases} \quad (3)$$

missä  $\gamma \in \mathbb{R}$  on ääriarvoindeksi (extreme value index, EVI).

Yleensä funktiolle  $G_\gamma(ax + b)$  käytetään käytännössä muotoa  $G_{\gamma, \mu, \sigma}(\frac{x + \mu}{\sigma})$ , jossa  $\mu > 0$  on lokaatioparametri ja  $\sigma \in \mathbb{R}$  on skaalaparametri. Ääriarvoindeksiä  $\gamma$  sanotaan myös muotoparametriksi.

Funktio  $G_\gamma(ax + b)$  on yleistetty ääriarvojakauma, joka voidaan jakaa kolmeen jakaumaperheeseen sen mukaan, mitä arvoja ääriarvoindeksi  $\gamma$  saa. Näitä jakaumia kutsutaan ääriarvojakaumiksi. Käytetään merkintää  $\alpha = 1/\gamma$  kun  $\gamma \neq 0$  ja  $\alpha = 0$  kun  $\gamma = 0$ . Tarkastellaan seuraavaksi funktion  $G_\alpha(x)$  jakaumia eri parametrin  $\alpha$  arvoilla. Kuvassa 1 on esimerkkejä ääriarvojakaumien tiheysfunktioista ja kuvassa 2 niiden kertymäfunktioista.

Kun  $\alpha < 0$ , parametrisoidaan  $G_\alpha(-\alpha(1 + x))$  ja yhtälöstä (3) saadaan *Weibull*-jakauma:

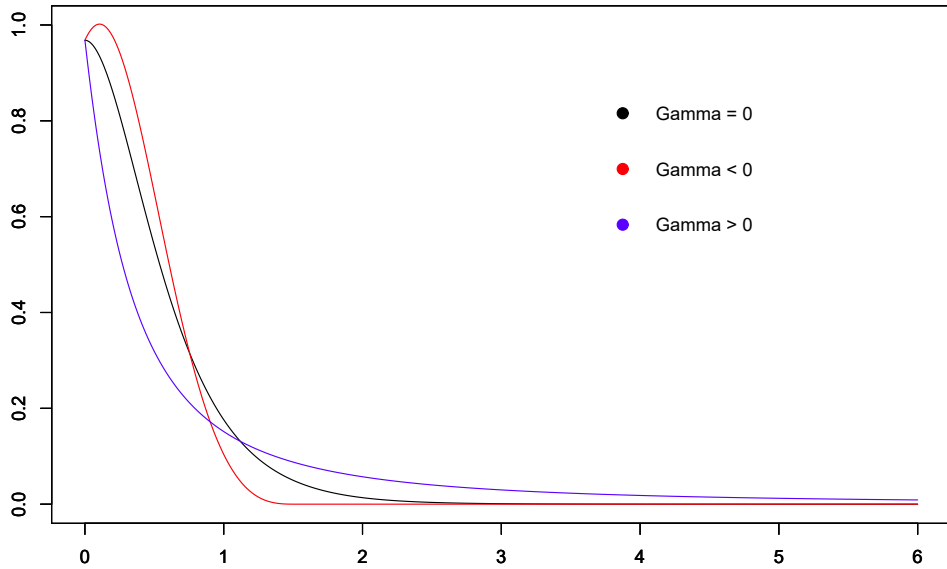
$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & x < 0 \\ 1, & x \geq 0. \end{cases} \quad (4)$$

Kun  $\alpha > 0$ ,  $G_\alpha(\alpha(x - 1))$  on *Fréchet*-jakauma:

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \end{cases} \quad (5)$$

ja kun  $\gamma = 0$  eli  $\alpha = 0$  niin  $G_\alpha(x)$  on *Gumbel*-jakauma

$$\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}. \quad (6)$$

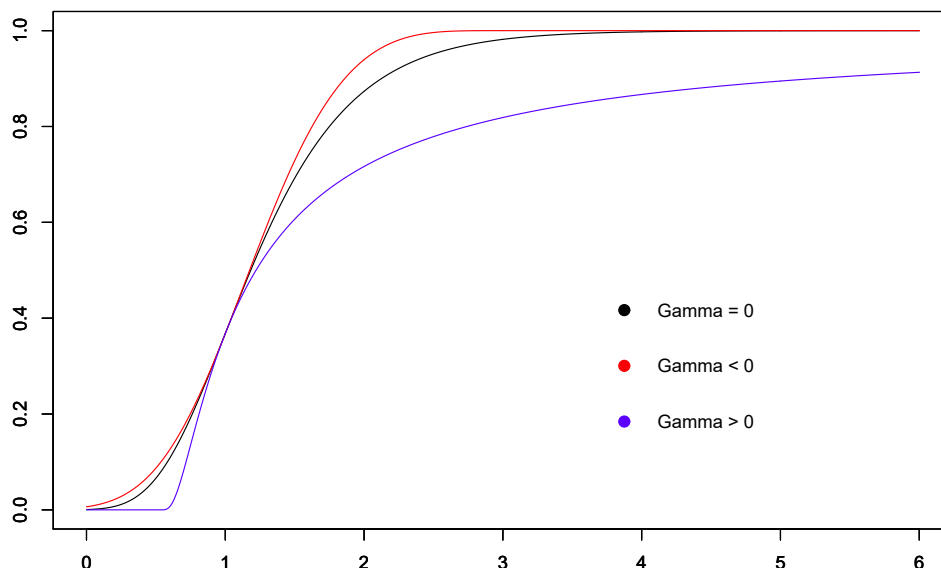


Kuva 1: Esimerkkejä yleistetyn ääriarvojakaumaperheen tiheysfunktioista eri parametrin  $\gamma$  arvoilla.

On huomionarvoista, että lause 3.1 ei takaa otosmaksimien  $M_n$  rajajakauman  $G(x)$  olemassaoloa. Se kuitenkin takaa, että jos sopivat normeerusvakiot löytyvät ja  $G(x)$  on olemassa, se on välttämättä lauseen määräämää muotoa riippumatta samoin jakautuneiden ja riippumattomien satunnaismuuttujien jakaumasta  $F$ . Käytännössä  $G(x)$  on melkein aina olemassa. Kaikilla tavanomaisilla jatkuvilla jakaumilla on olemassa ääriarvojakauma ja sellaisia ovat myös vakuutus- ja finanssialalla käytettävät funktiot.

Jos rajajakauma  $G(x)$  on olemassa, ja  $G(x) = G_\gamma(x)$  jollain  $\gamma \in \mathbb{R}$ , sanotaan, että satunnaismuuttujien jakauma  $F$  kuuluu jakauman  $G_\gamma(x)$  vaikutuspiiriin maksimin suhteen. Jakauman  $F$  muoto vaikuttaa siihen, min-





Kuva 2: Esimerkkejä yleistetyn ääriarvojakaumaperheen kertymäfunktioista eri parametrin  $\gamma$  arvoilla.

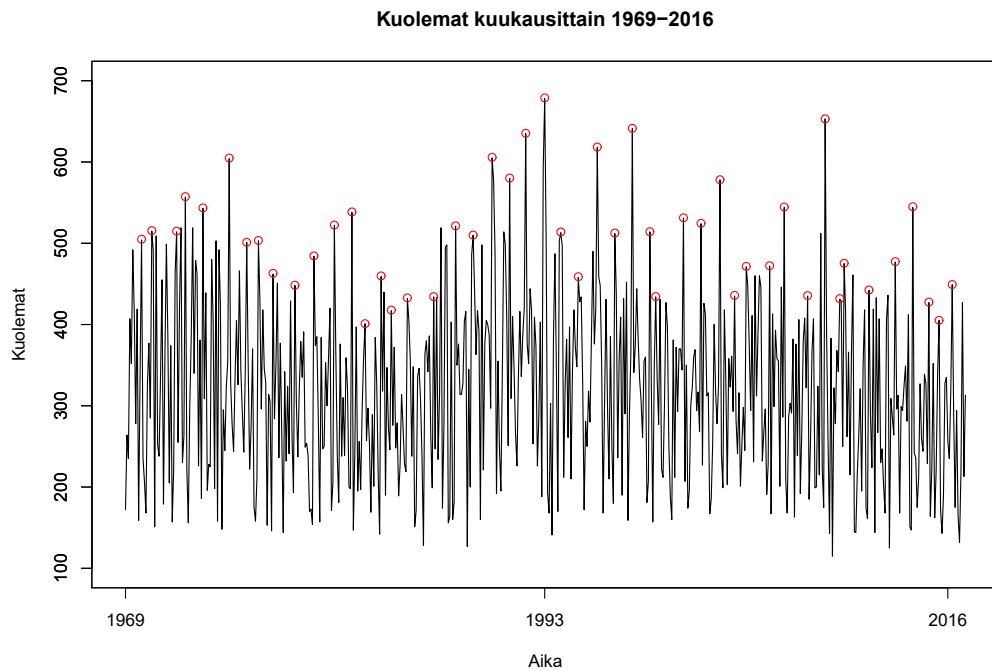
kä ääriarvojakauman vaikutuspiiriin se kuuluu maksimin suhteen. Esimerkiksi Fréchet-jakauman vaikutuspiiriin kuuluu paksuhäntäisiä jakaumia, sillä Fréchet-jakaumalla on ääretön oikea päätepiste ja ääriarvojakaumista paksuin häntä.

*Blokkimaksimimenetelmä* (method of block maxima) on ääriarvoteoriaa soveltava estimointimenetelmä parametreille  $\mu, \sigma$  ja  $\gamma$ . Se hyödyntää ääriarvojakaumien määrittelmää. Blokkimaksimimenetelmä perustuu ajatukseen, että suurilla parametrin  $n$  arvoilla saadaan ääriarvojakaumalle tarpeeksi hyvä estimaatti, jota voidaan käyttää äärimmäisten havaintojen ennustamiseen. Suurilla parametrin  $n$  arvoilla

$$\mathbb{P}(M_n \leq x) = F^n\left(\frac{x + \mu}{\sigma}\right) \approx G_\gamma(x). \quad (7)$$

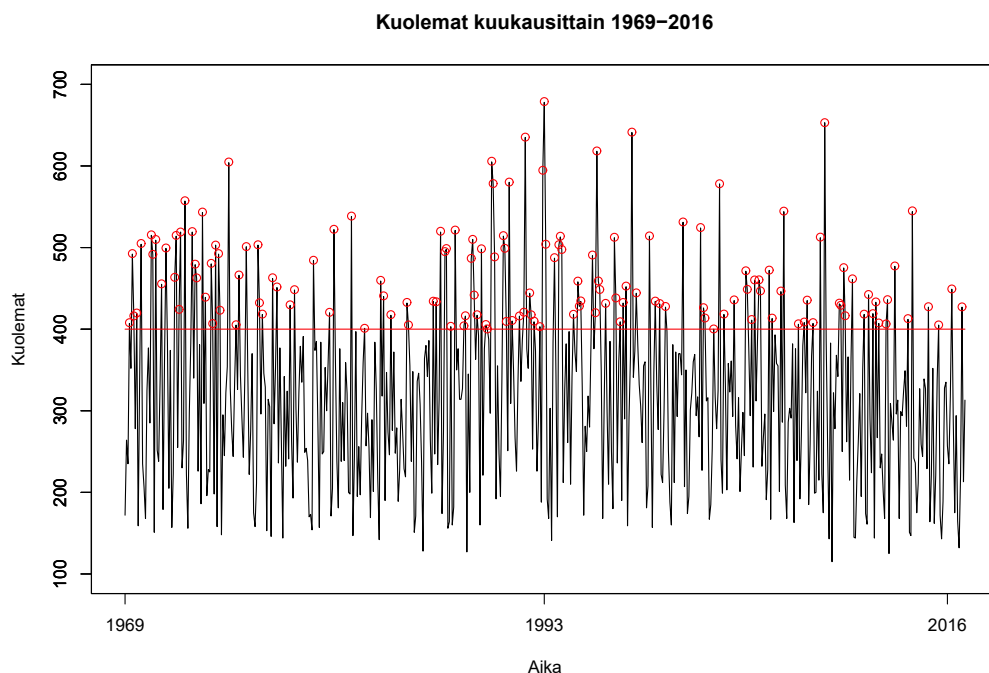
Tämän hyödyntämiseksi aineisto jaetaan  $n$  havainnon blokkeihin, joista poimitaan maksimiarvo  $M_n$ . Päivittäisiä havaintoja tarkasteltaessa hyvä ko-

ko blokeille voisi olla vuosi, jolloin tarkasteltaisiin vuosittaisia otosmaksimeita. Näihin maksimeihin sovitetaan yleistetty ääriarvojakauma estimoimalla muotoparametria  $\mu$ , skaalausparametria  $\sigma$  ja ääriarvoindeksiä  $\gamma$  tilastollisia menetelmiä käyttäen. Tästä jakaumasta saadaan ääriarvojen esiintymistodennäköisyyksille estimaatit. Kuvassa 3 on aikasarjana esitetty kuukausittaisia havaintoja ja niistä vuoden maksimiarvot on korostettu punaisilla ympyröillä. Näihin vuosittaisiin maksimiarvoihin voidaan sitten sovittaa yleistetty ääriarvojakaumaa.



Kuva 3: Aikasarja kuukausittaisista kuolemien määrästä, josta punaisella ympyröitynä blokkimaksimit. Blokkina on yksi vuosi. Aikasarja on luotu Tilastokeskuksen vuosittaisten kuolemien tilastosta [14] simuloimalla kuukausijaksotusta tasajakaumasta  $U(0, 1)$  poimittujen satunnaislukujen avulla. Tilastoa käytetään myöhemmin osiossa 4.2 *Mallinnus* ja aineiston simulointia kuukausitasolle on esitelty tarkemmin osiossa 4.2.1 *Aineiston esittely*.

### 3.3.2 Yleistetty Pareto-jakauma ja ylitemenetelmä



Kuva 4: Aikasarja kuukausittaisista kuolemien määrästä, josta punaisella viivalla kynnyisarvo  $u = 400$  ja punaisella ympyröitynä sen ylittävät havainnot. Tilastoa käytetään myöhemmin osiossa *4.2 Mallinnus*.

Blokkimaksimimenetelmä ei ole optimaalinen kaikissa sovelluksissa. Koska se ottaa huomioon vain yhden ääriarvon jokaisesta blokista, voi sitä käyttäessä jäädä monta äärimmäiseksi luokiteltavaa havaintoa hyödyntämättä mallinnuksessa. Käytännön sovelluksissa havainnot eivät yleensä ole täysin riippumattomia, vaan äärimmäiset arvot esimerkiksi meriveden korkeuksille havaitaan usein lähekkäin luonnonilmiöiden seurauksena. Etenkin tällöin jää sellaisilta poikkeuksellisilta vuosilta paljon aineistoa käyttämättä, joilla ilmenee monta äärimmäistä havaintoa. Tästä johtuen määritellään kynnyisarvo, jonka ylittäviä havaintoja pidetään äärimmäisinä. Esitellään seuraavaksi Colesin (2001) lause lähteestä [1]. Lauseen todistus löytyy lähdeeteoksesta.

**Lause 3.2** (Yleistetty Pareto-jakauma). Olkoon  $(X_i), i = 1, \dots, n$  jono riippumattomia ja samoin jakautuneita satunnaismuuttujia, joilla on yhteiskertymäfunktio  $F$  ja

$$M_n = \max\{X_1, \dots, X_n\}. \quad (8)$$

Olkoon  $X$  eräs näistä muuttujista ja oletetaan että  $F$  toteuttaa edellisessä osiossa esitellyn yhtälön (7), eli maksimien rajajakauma on yleistetty ääriarvojakauma joillain parametreillä  $\gamma, \mu$  ja  $\sigma$ . Merkitään kynnyksarvoa kirjaimella  $u$ . Tällöin ehdolla  $X > u$  on kynnyksen ylitteen  $(X - u)$  jakauma tarpeeksi suurella kynnyksarvolla  $u$  likimäärin

$$H(y) = \begin{cases} 1 - (1 + \gamma \frac{y}{\sigma})^{-1/\gamma}, & \gamma \neq 0 \\ 1 - e^{-y/\bar{\sigma}}, & \gamma = 0 \end{cases} \quad (9)$$

kun  $y > 0, (1 + \gamma \frac{y}{\sigma}) > 0$  ja  $\bar{\sigma} = \sigma + \gamma(u - \mu)$  missä  $\mu \in \mathbb{R}$ .

Lauseesta seuraa, että jos blokkimaksimeiden jakauma on likimäärin yleistetty ääriarvojakauma  $G$ , niin kynnyksarvon ylitteille jakauma on vastaavasti likimäärin yleistetty Pareto-jakauma  $H$ . Huomataan myös, että Pareto-jakauman  $H$  parametrit ovat itseasiassa yhteydessä jakauman  $G$  parametreihin. Niillä on nimittäin sama muotoparametri  $\gamma$  ja skaalaparametrit ovat myös yhteydessä  $\bar{\sigma} = \sigma + \gamma(u - \mu)$  kuten lauseessa mainittiin. Jatkossa käytetään Pareto-jakauman skaalaparametrille  $\bar{\sigma}$  yksinkertaistuksen vuoksi vain merkintää  $\sigma$ , mutta muistetaan, että se ei ole täysin sama parametri kuin yleistetylle ääriarvojakaumalle.

Esitellään lauseen 3.2 pohjalta yleistetty Pareto-jakauma  $H_{\gamma,\sigma}(x)$  ja määrittellään se seuraavasti

$$H_{\gamma,\sigma}(x) = \begin{cases} 1 - (1 + \gamma \frac{x}{\sigma})^{-1/\gamma}, & \gamma \neq 0 \\ 1 - e^{-x/\sigma}, & \gamma = 0, \end{cases} \quad (10)$$

missä  $\sigma > 0$  on skaalaparametri. Lisäksi voidaan määritellä sijaintiparametri  $\rho \in \mathbb{R}$ , jolloin  $H_{\gamma,\rho,\sigma}(x) = H_{\gamma,\sigma}(x - \rho)$ . Jakauma  $H_{\gamma,\sigma}(x)$  on määritelty alueella  $x \in \mathbb{R}^+$  kun  $\gamma \geq 0$  ja  $x \in [0, -\sigma/\gamma[$  kun  $\gamma < 0$ . Itseasiassa yleistetty Pareto -jakaumakin voidaan jakaa kolmeen alaluokkaan, kuten yleistetty

ääriarvojakaumakin. Kun  $\gamma > 0$ ,  $H_{\gamma,\sigma}$  on tavallinen Pareto-jakauma  $F_{\alpha,\kappa}$ , missä  $\alpha = 1/\gamma$  ja  $\kappa = \sigma/\gamma$ . Kun  $\gamma = 0$ ,  $H_{\gamma,\sigma}$  on eksponenttijakauma  $F_\lambda$ , missä  $\lambda = 1/\sigma$ . Kun  $\gamma < 0$ , jakauman oikea päätepiste  $x_F$  on äärellinen, ja  $x_F = -\sigma/\gamma$ .

Toisinaan on hyödyllistä tarkastella tason  $u$  ylitteiden suuruutta tarkemmin sen sijaan, että tarkasteltaisiin tason ylittäviä havaintoja. Leppisaaren (2013) [9] mukaan tason  $u$  ylitteiden suuruuksille voidaan määritellä ehdollinen odotusarvofunktio ehdolla, että taso  $u$  ylitetään. Tämä odotusarvo ei välttämättä ole olemassa, mutta jos on, se on muotoa

$$e(u) = \mathbb{E}(X - u | X > u). \quad (11)$$

Ylitteen odotusarvofunktio on yleistetylle Pareto-jakaumalle

$$e(u) = \frac{\sigma(u)}{1-\gamma} = \frac{\sigma + \gamma u}{1-\gamma}, \quad (12)$$

missä  $u \in [0, \infty)$  kun  $\gamma \in [0, 1)$  ja  $u \in [0, \frac{-\sigma}{\gamma}]$  kun  $\gamma < 0$ . Siis myös ylitteen odotusarvofunktio on ylitejakauman skaalaparametrin tavoin lineaarinen funktio, joka kasvaa kynnystason  $u$  suhteen.

Käytännön sovelluksia varten on tarpeen huomioida muutamia yleistetyn Pareto-jakauman sekä ylitejakauman ominaisuuksia. Ensinnäkin yleistetylle Pareto-jakaumalle  $\lim_{\gamma \rightarrow 0} H_{\gamma,\sigma}(x) = H_{0,\sigma}(x)$ , eli parametrisointi on kiinteällä  $x$  jatkuva parametrin  $\gamma$  suhteen. Tätä ominaisuutta tarvitaan esimerkiksi suurimman uskottavuuden estimointimenetelmää käytettäessä. Lisäksi kuten aiemmin todettiin, on ylitteen odotusarvofunktio  $e(u)$  kynnystason  $u$  suhteen lineaarinen funktio. Tämä ominaisuus säilyy myös korkeampien kynnysarvojen ylitteitä tarkasteltaessa kuin  $u$ . Kynnysarvolla  $v > u$  on

$$e(v) = \frac{\sigma + \gamma(v - u)}{1 - \gamma} = \frac{\gamma v}{1 - \gamma} + \frac{\sigma - \gamma u}{1 - \gamma}. \quad (13)$$

Tämän ominaisuuden ansiosta käytännön sovelluksissa voidaan löytää sopiva kynnysarvo  $u$  seuraavasti. Ensin täytyy muodostaa odotusarvolle estimaatti arvon  $u$  ylittävien havaintojen avulla, eli laskemalla otoskeskiarvo havaituista ylitteistä. Jos ylitteet noudattavat yleistettyä Pareto-jakaumaa, niin yhtälön (12) mukaan otoskeskiarvojen tulisi olla lineaarisia eri kynnysarvon

$u$  arvoilla. Piirtämällä otoskeskiarvon kuvaaja eri kynnsarvoja vasten ja esimällä kuvaajasta alue, josta alkaen funktio on lähes lineaarinen, voidaan täten löytää sopiva kynnsarvo.

Coles (2001) [1] on määritellyt kynnsarvon löytämiseksi toisenkin menetelmän. Käytetään kynnsarvolle  $u > u_0$  sen yleistetyn Pareto-jakauman skaalaparametrille merkintää  $\sigma_u$ . Nyt aiemmin esitetyn yhtälön  $\bar{\sigma} = \sigma + \gamma(u - \mu)$  perusteella saadaan

$$\sigma_u = \sigma_{u_0} + \gamma(u - u_0). \quad (14)$$

Tämän mukaan skaalaparametri muuttuu kynnsarvon mukaan kaikissa muissa tapauksissa paitsi kun  $\gamma = 0$ . Käytetään uudelleenparametrisointia ja saadaan *muunnettu skaalaparametri*

$$\sigma^* = \sigma_u - \gamma\mu \quad (15)$$

joka on vakio parametrin  $u$  suhteen. Nyt jos  $u$  on sopiva kynnsarvo eli sen ylitteet noudattavat yleistettyä Pareto-jakaumaa, estimaatit  $\gamma$  ja  $\sigma^*$  pysyvät vakioina. Eli voidaan piirtää muunnetun skaalaparametrin ja parametrin  $\gamma$  kuvaajat eri kynnsarvoilla  $u$  ja valita sellainen arvo  $u$  josta lähtien  $\gamma$  ja  $\sigma^*$  pysyvät luottamusväleineen mahdollisimman stabiilena. Tätä menetelmää on käytetty myöhemmin kappaleessa *4.2 Mallinnus*.

## 4 Ääriarvoteoria yhdistettynä Value-at-Riskiin

Luvussa *2.4.1 Parametrinen Value-at-Risk* kerrottiin VaR:in määrittelemisestä todennäköisyysjakauman avulla. VaR:in estimaatista saadaan luotettavampi, jos ääriarvoteoriaa hyödynnetään todennäköisyysjakauman hännän mallintamiseen. Tarkastellaan VaR:in ja ääriarvoteorian yhdistämistä ensin yleistetyn ääriarvojakauman ja sitten blokkimaksimimenetelmän näkökulmasta. Käytetään pohjana Leppisaaren (2013) [9] käyttämiä määritelmiä.

### 4.1 Toistumisjakso ja toistumistaso

Käytännön sovelluksissa ääri-ilmiöitä tarkasteltaessa käytetään usein termejä *toistumisjakso* ja *toistumistaso*. Seuraavissa määritelmissä oletetaan ää-

rimmäisiä arvoja  $X_i$  luovat prosessit stationaarisiksi. Epästationaarisilla prosesseilla toistumisjakso ja toistumistaso eivät ole vakioita yli ajan. Niille tapahtumien todennäköisyydet muuttuvat ajassa havaintojen jakaumien muuttuessa. Lisäksi oletetaan, että toistumisjaksoa voidaan ajatella tässä joko ajanjaksona tai havaintojen lukumääränä, sillä laskentatapa on molemmissa tapauksissa sama. Merkitään toistumisjaksoa  $t_u$  ja määritellään se jonolle satunnaismuuttujia  $(X_i)$ , joilla on yhteiskertymäfunktio  $F$ . Kynnystasoa  $u$  suurempien tapahtumien  $X_i > u$  toistumisjakso on

$$t_u = \frac{1}{1 - F(u)}. \quad (16)$$

Tämä tulkitaan niin, että ajassa  $t_u$  tai havainnossa  $t_u$  sattuu keskimäärin yksi tason ylitys. Toistumisjakso siis kertoo, minkä verran aikaa tai vastaavasti havaintoja on keskimäärin kahden kynnystason ylittävän havainnon välillä. On huomionarvoista, että joitain ilmiöitä kuten meriveden korkeutta tarkasteltaessa on oletettavaa, että useampi äärimmäinen havainto esiintyy peräkkäin, vaikka toistumisjakson mukaan äärimmäisiä arvoja havaitaan kerran  $t_u$  vuodessa.

Määritellään seuraavaksi toistumistaso, ja käytetään siitä merkintää  $x_t$ . Nyt jonolle satunnaismuuttujia  $(X_i)$  yhteisellä kertymäfunktiolla  $F$  määritellään toistumisjaksoa  $t$  vastaava toistumistaso

$$x_t = q_{1-1/t}(F) = F^{\leftarrow} \left( 1 - \frac{1}{t} \right). \quad (17)$$

Siis toistumistaso on jakauman  $F$   $(1 - 1/t)$ -kvantiili, ja sen tulkinta on taso, joka ylitetään keskimäärin kerran jaksossa  $t$ . Suoraan määritelmästä seuraa, että jakson  $t_u$  toistumistaso on  $u$ . Toistumistason määritelmästä huomataan, että se on itseasiassa  $\text{VaR}_{1-\frac{1}{t}}$ .

#### 4.1.1 Yleistetty ääriarvojakauma

Ääriarvojakaumaan  $\text{VaR}_{1-\frac{1}{t}}$  saadaan yhdistettyä hakemalla toistumistasolle eksplisiittistä esitystä yleistetyn ääriarvojakauman avulla. Esitys perustuu Colesin (2001) teokseen [1]. Avuksi tarvitaan yleistetyn ääriarvojakauman määritelmä (3) kappaleesta *Ääriarvojakaumat ja blokkimaksimimenetelmä*.

Käytetään jakaumaa parametrilla  $\frac{(1-\frac{1}{t})-\mu}{\sigma}$ , missä  $\mu \in \mathbb{R}$  ja  $\sigma < 0$ . Se voidaan kirjoittaa muodossa

$$G_{\gamma,\mu,\sigma}\left(1 - \frac{1}{t}\right) = \begin{cases} \exp\left(-\left(1 + \gamma\frac{(1-\frac{1}{t})-\mu}{\sigma}\right)^{-1/\gamma}\right), & \gamma \neq 0 \\ \exp\left(-e^{\frac{(1-\frac{1}{t})-\mu}{\sigma}}\right), & \gamma = 0. \end{cases} \quad (18)$$

Ottamalla tästä käänteisfunktio saadaan kaivattu eksplisiittinen esitys toistumistasolle  $x_t$  eli  $\text{VaR}_{1-\frac{1}{t}}$

$$x_t = G_{\gamma,\mu,\sigma}^{-1}\left(1 - \frac{1}{t}\right) = \begin{cases} \mu - \frac{\sigma}{\gamma} \left(1 - \{-\ln(1 - \frac{1}{t})\}^{-\gamma}\right), & \gamma \neq 0 \\ \mu - \sigma \ln\{-\ln(1 - \frac{1}{t})\}, & \gamma = 0. \end{cases} \quad (19)$$

Estimoimalla parametrit  $\hat{\gamma}$ ,  $\hat{\rho}$  ja  $\hat{\sigma}$  ja sijoittamalla ne yhtälöön saadaan toistumistasolle estimaatti  $\hat{x}_t$ . Käytännön sovelluksissa on usein hyödyllistä tarkastella toistumisjaksoa ja -tasoa graafisesti. Sopivalla skaalauksella toistumistasolle voidaan muodostaa kuvaaja toistumisjakson funktiona. Toistumisjakson kuvaaja on helppo muodostaa tästä kääntämällä vain akselit toisinpäin, kuten kuvassa 5. Piirtämistä voi helpottaa sijoittamalla toistumistason funktion  $y_t = -\ln(1 - \frac{1}{t})$ , jolloin yhtälö voidaan kirjoittaa yksinkertaisemmin muodossa

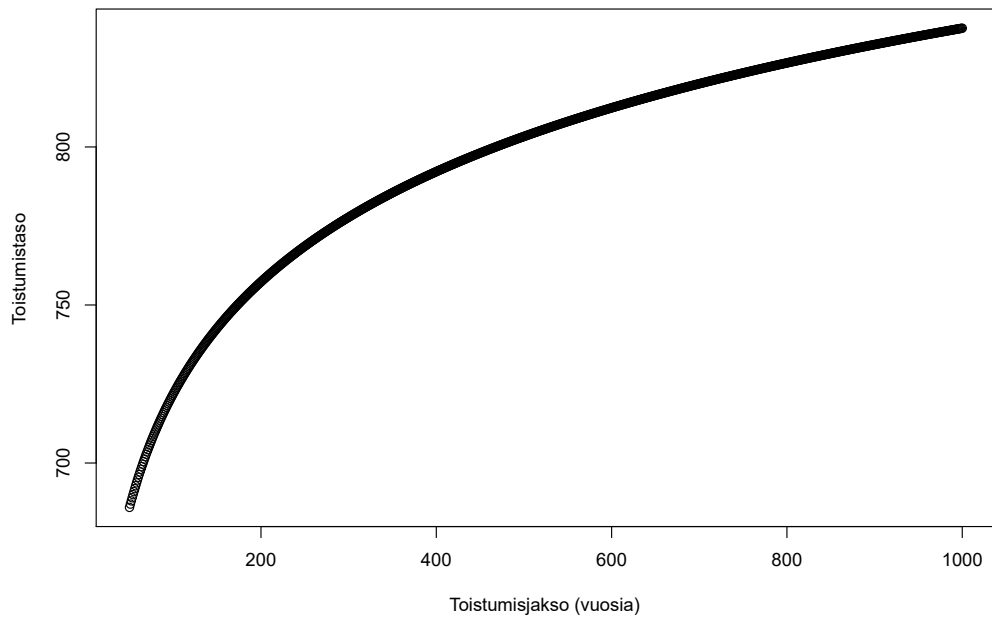
$$x_t = G_{\gamma,\mu,\sigma}^{-1}\left(1 - \frac{1}{t}\right) = \begin{cases} \mu - \frac{\sigma}{\gamma} (1 - y_t^{-\gamma}), & \gamma \neq 0 \\ \mu - \sigma \ln(y_t), & \gamma = 0. \end{cases} \quad (20)$$

Esimerkiksi osiossa *4.2 Mallinnus* käytetyssä *R*-ohjelmointikielen paketissa *evd* mallin sopivuutta tarkasteleva funktio käyttää Stephensonin (2002) [13] mukaan tätä muotoa toistumistason kuvaajan piirtämiseen.

#### 4.1.2 Yleistetty Pareto-jakauma

Yleistetyn Pareto-jakauman kanssa, eli käytettäessä ylitemenetelmää, käsitellään ehdollista todennäköisyyttä. Tässä kappaleessa on käytetty pohjana Colesin (2001) [1] esitystapaa. Oletetaan että muuttujalle  $X$  tason  $u$  ylitteet noudattavat yleistettyä Pareto-jakaumaa parametreilla  $\gamma$  ja  $\mu$ . Merki-





Kuva 5: Esimerkkikuva toistumisjakson ja toistumistason graafisesta tarkastelusta.

tään  $p_u = \mathbb{P}(X > u)$ . Saadaan

$$\mathbb{P}(X > x | X > u) = \left(1 + \gamma \left(\frac{x - u}{\sigma}\right)\right)^{-1/\sigma}, x > u \quad (21)$$

josta saadaan edelleen

$$\mathbb{P}(X > x) = p_u \left(1 + \gamma \left(\frac{x - u}{\sigma}\right)\right)^{-1/\sigma}, x > u. \quad (22)$$

Yleistetylle Pareto-jakaumalle saadaan tästä muotoiltua vastaavalla menetelyllä kuin yleistetylle ääriarvojakaumallekin toistumistaso  $x_t$  eli  $\text{VaR}_{1-\frac{1}{t}}$ . Määritellään toistumistaso  $x_t$  vastaavasti kuin yleistetyn ääriarvojakauman tapauksessa tasoksi, jota vastaa toistumisjakso  $t$ . Siis  $x_t$  on taso joka ylittää kerran  $t$ :ssä havainnossa. Tämä voidaan kirjoittaa yhtälön (22) mukaisesti muodossa

$$\frac{1}{t} = \begin{cases} p_t \left(1 + \gamma \left(\frac{x_t - u}{\sigma}\right)\right)^{-1/\sigma} & \gamma \neq 0 \\ p_t e^{-\frac{x_t - u}{\sigma}}, & \gamma = 0 \end{cases} \quad (23)$$

josta kummastakin tapauksesta saadaan edelleen ratkaistua  $x_t$

$$x_t = \begin{cases} u + \frac{\sigma}{\gamma} ((tp_u)^\gamma - 1), & \gamma \neq 0 \\ u + \sigma \ln(tp_u), & \gamma = 0. \end{cases} \quad (24)$$

Käytännön sovelluksissa halutaan usein tarkastella vuosiksi skaalattuja toistumisjaksoja. Käytetään seuraavaksi vuosien määrästä eli vuosiksi skaalattusta toistumisjaksosta merkintää  $N$  ja havaintojen määrästä vuodessa merkintää  $n_y$ . Tällöin  $t = Nn_y$  ja kaavaan (24) sijoittamalla saadaan  $N$  vuoden toistumistasoksi

$$x_N = \begin{cases} u + \frac{\sigma}{\gamma} ((Nn_y p_u)^\gamma - 1), & \gamma \neq 0 \\ u + \sigma \ln(Nn_y p_u), & \gamma = 0. \end{cases} \quad (25)$$

Vastaavasti kuin yleistetyn ääriarvojakauman kanssa toistumistasolle laskeaan estimaatti  $\hat{x}_n$  estimoimalla parametrit  $\hat{\gamma}$  ja  $\hat{\sigma}$  ja sijoittamalla ne yhtälöön

(25). Luonnollisena estimaattina todennäköisyydelle  $p_u$  toimii tason  $u$  ylittävien havaintojen osuus. Eli  $\hat{p}_u = \frac{n_u}{n_m}$  missä  $n_u$  on kynnyksen  $u$  ylittävien havaintojen lukumäärä ja  $n_m$  on kaikkien havaintojen lukumäärä.

## 4.2 Mallinnus

Tässä osiossa esitellään miten blokkimaksimimenetelmää ja ylitemenetelmää sovelletaan käytännössä. Esitellään ensin aineisto, jonka jälkeen luodaan mallit ensin blokkimaksimimenetelmällä ja sitten ylitemenetelmällä. Ylitemenetelmälle tehdään kaksi mallia eri kynnyksarvojen  $u$  antamien mallien vertailemiseksi.

Mallinnukseen käytetään  $R$ -ohjelmointikieltä. Ääriarvojen mallintamiseen valitaan Stephensonin (2002) paketti *evd* [13], koska sen avulla voidaan soveltaa sekä blokkimaksimimenetelmää että ylitemenetelmää. Paketti valitaan myös sen hyvien piirtokomentojen vuoksi jotta graafinen tarkastelu on mahdollisimman helppoa.

### 4.2.1 Aineiston esittely

Aineisto on valittu sillä perusteella, että se kuvaisi jotain todellista vakuutusalaan liittyvää ilmiötä ja olisi peräisin virallisesta lähteestä. Aineistoksi valikoidaan Tilastokeskuksen julkaisema tilasto kuolemansyistä [14], ja tilastosta käytetään kuolleiden kokonaisuuttua. Aineiston havainnot on piirretty kuvassa 6. Kyseistä tilastoa voitaisi periaatteessa hyödyntää esimerkiksi henkivakuutusten maksimikorvausarvion selvittämiseen. Tilasto sisältää vuodesta 1969 vuoteen 2016 asti vuosittaisen kuolleiden henkilöiden lukumäärän. Koska havaintoja on niin vähän, 48 kappaletta, simuloidaan aineisto vielä vuositasolta kuukausitasolle. Tämä tehdään  $R$ -ohjelmointikielellä. Halutaan luoda kuukausijaottelu joka sisältää satunnaisuutta, mutta on silti lähellä kahdestoistaosaa kunkin vuoden kuolemista.

Simuloitu aineisto luodaan seuraavasti. Jokaisen vuoden osalta kuhunkin kuukauteen asetetaan yksi kerrallaan kuolemien määrä kaavalla  $2r \frac{k}{12}$ , missä  $r$  on tasajakaumasta  $U(0, 1)$  poimittu satunnaisluku ja  $k$  on kyseisen vuoden kuolemien määrä. Rajoitetaan lisäksi, että kuukausissa pitää olla vähintään

$0.55\frac{k}{12}$  kuolemaa, koska oletetaan että ihmisiä kuolee melko tasaisesti jokaisena kuukautena. Kuhunkin kuukauteen asetetaan siis kuolemien määrä, joka on  $0.55 - 2$ -kertainen verrattuna tasaisesti jaettuun kuukausittaiseen kuolinmäärään kyseisenä vuonna.

Lisäksi halutaan että kuukausille simuloitujen kuolemien määrät vastaavat vuositasolla mahdollisimman hyvin alkuperäistä aineistoa. Siksi käytetään ehtoa, että jos simuloinnin jälkeen vuodessa kuukausien yhteenlaskettu kuolleiden määrä on pienempi kuin 95% Tilastokeskuksen tilaston kyseisestä vuodesta, lisätään erotus joulukuuhun. Näin joinain vuosina joulukuuhun tulee helposti vuoden korkein kuolinmäärä. Se ei kuitenkaan haittaa, koska nyt ei olla kiinnostuneita kuukausien järjestyksestä. Toisaalta simuloitaessa vuosittainen kuolleiden määrä voi tulla täyteen jo ennen kuin jokaiseen kuukauteen on asetettu luku. Tällöin pienennetään kaikkia lukuja 5% ja jatketaan kuolemien simuloimista.

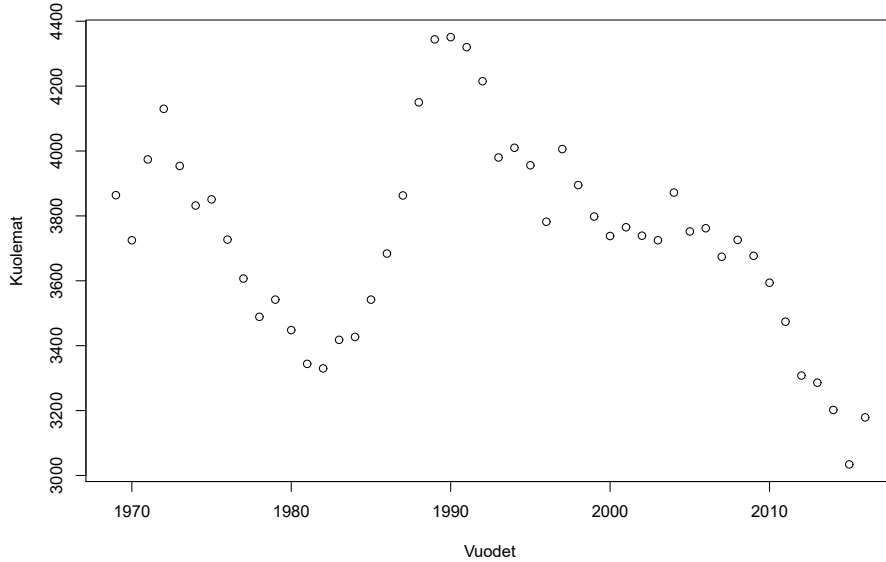
Edellisten ehtojen mukaan toimittaessa useimpien vuosien kuolleiden määrä simuloidussa aineistossa poikkeaa hieman alkuperäisestä. Ero alkuperäiseen aineistoon on käytetyllä simuloidulla aineistolla välillä  $[-2\%, +10\%]$ .

Simuloitu aineisto vaikuttaa kuvan 4 perusteella stationaariselta. Voimme siis käyttää aiemmissa kappaleissa esitettyjä menetelmiä toistumistason määrittämiseksi. Koska kuukausittainen jako on keinotekoinen, pitää mallinnuksen antamia tuloksia kuitenkin tarkastella tavallistakin kriittisemmin.

Vaikka kuolemien määrä on diskreetti muuttuja, kohdellaan sitä jatkossa mallinnuksen helpottamiseksi jatkuvana.

#### 4.2.2 Blokkimaksimimenetelmä

Sovelletaan aineistoon ensin blokkimaksimimenetelmää. Blokiksi luonnollinen valinta on vuosi, vaikka yksittäisessä blokissa on simuloitunkin aineiston tapauksessa suhteellisen vähän havaintoja. Aiemmin esitelystä kuvasta 3 nähdään vuosittaiset maksimihavainnot kuukausittaisten kuolemien aikasarjassa punaisina ympyröinä. Erityisesti simuloimattomalla aineistolla kuolevuus voi olla korkealla tasolla useana peräkkäisenä kuukautena esimerkiksi sodan tai epidemian takia. Siksi olisi perusteltua käyttää todellisessa tilan-



Kuva 6: Tilastokeskuksen tilasto vuosittaisista kuolemista vuosilta 1969 – 2016 [14]. Tilastoa on käytetty kuukausitasolle simuloidun aineiston pohjana. Kuvassa 90-luvun lama näkyy korkeampana kuolleisuutena.

teessa pelkästään ylitemenetelmää ilmiön mallintamiseen. Nyt halutaan kuitenkin käyttää molempia menetelmiä.

Käytetään ääriarvojakauman mallintamiseen *R*-ohjelmointikielen pakettia *evd* [13]. Tarkoitus on siis löytää aineiston perusteella yleistetylle ääriarvojakaumalle

$$G_{\gamma,\mu,\sigma}(x) = \begin{cases} \exp\left(-\left(1 + \gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right), & 1 + \gamma\frac{x-\mu}{\sigma} > 0, \\ \exp\left(-e^{-\frac{x-\mu}{\sigma}}\right), & \gamma = 0 \end{cases} \quad (26)$$

estimaatit  $\hat{\gamma}$ ,  $\hat{\mu}$  ja  $\hat{\sigma}$ , jotka tuottavat estimaatin kuukausittaisten kuolemien ääriarvojen jakaumalle. Estimaattien löytämiseksi käytetään paketin *evd* funktiota *fgev*. Kyseinen funktio käyttää Stephensonin (2002) [13] mukaan suurimman uskottavuuden menetelmää (maximum likelihood estimation, MLE). MLE-menetelmän uskottavuusfunktion maksimoimiseksi funktio käyttää numeerista *Nelder-Mead* -optimointimenetelmää.

Mallin tuottamien estimaattien luottamusvälien määrittämiseksi käytetään paketin *evd* funktiota *profile* joka laskee luottamusvälit *logprofiliskottavuusmenetelmää* käyttäen. Menetelmässä maksimoidaan loguskottavuusfunktiota kullekin parametrille kaikkien muiden parametrien suhteen. Valitaan luottamusväliksi 99.5%. Kun piirretään funktion *profile* tuloste saadaan menetelmän tuottamat 99.5%:n luottamusvälit katkoviivan ja logprofiliskottavuusfunktion leikkauspisteissä.

Saadaan lokaatioparametriksi  $\hat{\mu} = 476.16$ , skaalaparametriksi  $\hat{\sigma} = 55.73$  ja muotoparametriksi  $\hat{\gamma} = -0.02$ . Kyseessä on siis Weibull-jakauma eli jakauma on lyhythäntäinen. Kuvissa 7 ja 8 on parametrien luottamusvälit. Huomataan, että muotoparametrin estimaatin  $\hat{\gamma}$  luottamusväli sisältää nollan, joten tapausta että kuukausittaisten kuolemien määrän äärihavainnot olisivat Gumbel-jakautuneita ei voida sulkea pois.

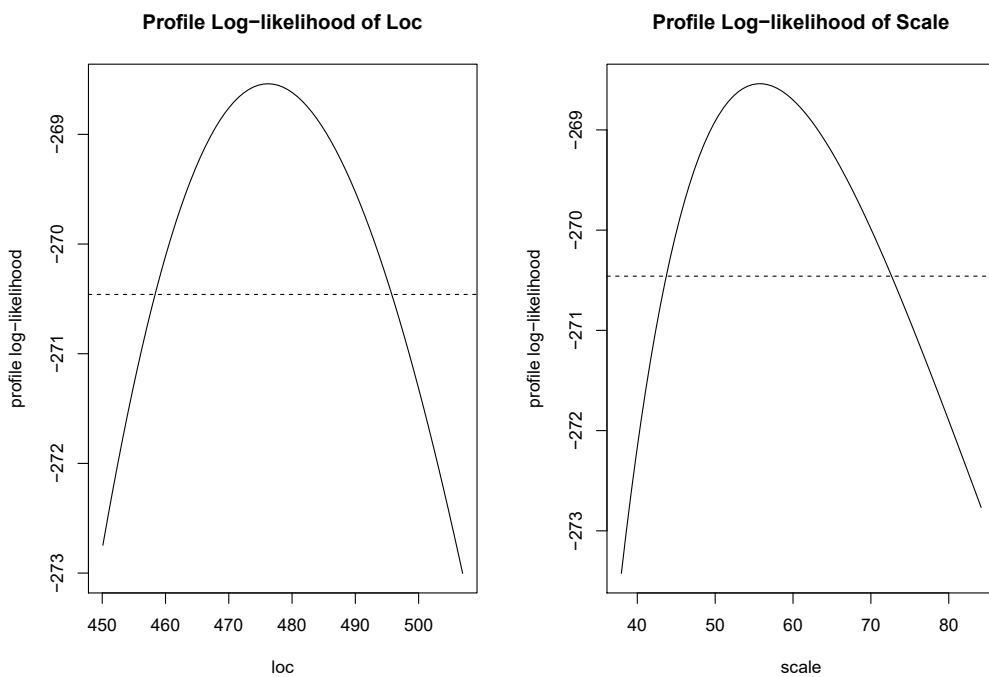
Tarkastellaan seuraavaksi saadun mallin sopivuutta aineistoon piirtämällä funktion *fgev* tuloste. Kuvassa 9 on mallin todennäköisyyskuvaaja (PP-plot) ja kvantiilikuvaaja (QQ-plot). Jos malli on hyvä, tulisi havaintojen osua diagonaalille molemmissa kuvaajissa. Koska havainnot osuvat melko hyvin diagonaalille ja luottamusvälien sisälle, eivät kuvaajat osoita mallin sopivuutta vastaan.

Kuvassa 10 taas on vasemmalla mallin tiheysfunktio ja aineiston empiirinen tiheysfunktio. Jos malli on hyvä, tulisi mallin antaman tiheysfunktion vastata empiiristä tiheysfunktiota. Huomataan, että malli mukaileekin sitä kohtuullisen hyvin. Erityisesti ottaen huomioon että havaintoja on melko vähän, vain 48.

Verrataan sitten mallin antamaa toistumistasokuvaajaa havaittuihin toistumistasoihin. Kuvassa 10 on oikealla piirretty toistumistasokuvaaja logaritmisella asteikolla, eli pisteet  $(\frac{-1}{\ln(p_t)}, x_t), t = 1, \dots, m$  missä  $m$  on aineiston havaintojen lukumäärä ja  $x_t$  on laskettu yhtälöstä (20) käyttäen tapausta  $\gamma \neq 0$ . Toistumistasoja tarkasteltaessa kaikki havainnot osuvat luottamusvälien sisään, ja sijoittuvat muutenkin kohtuullisen lähelle mallin antamia arvoja.

Tarkastellaan tuloksia vielä Value-at-Riskin näkökulmasta. Estimaattiksi parametrille  $\text{VaR}_{1-\frac{1}{t}}$  saadaan yhtälön (19) perusteella  $\mu -$

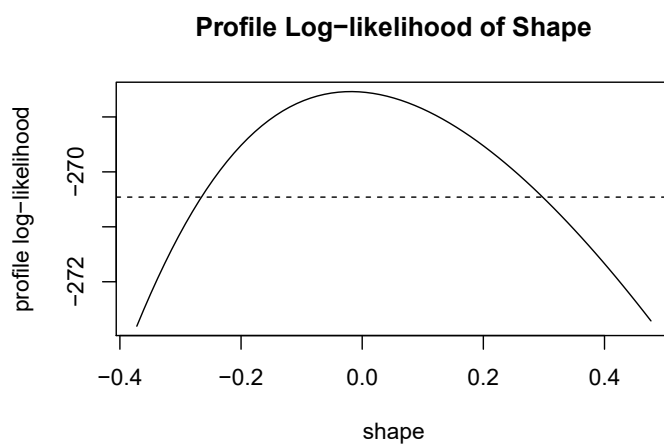
$\frac{\sigma}{\gamma} \left(1 - \left\{-\ln\left(1 - \frac{1}{t}\right)\right\}^{-\gamma}\right)$ . Halutaan tarkastella parametria niin että ajanjaksona on vuosi, joten käytetään aiemmin esiteltyä kaavaa  $t = Nn_y$ . Nyt sijoitetaan kaavaan parametri  $t$  sekä saadut estimaatit  $\hat{\mu} = 476.16$ ,  $\hat{\sigma} = 55.73$  ja  $\hat{\gamma} = -0.02$ . Estimaateiksi saadaan lähimpään kokonaislukuun pyöristämällä esimerkiksi  $\text{VaR}_{0.995} = 880$  ja  $\text{VaR}_{0.999} = 957$ . Kuvassa 11 on piirretty estimaatin  $\text{VaR}_{1-\frac{1}{t}}$  kuvaaja arvoilla  $\frac{1}{t} = \left(\frac{1}{100}, \frac{1}{101}, \dots, \frac{1}{1000}\right)$ .



Kuva 7: Blokkimaksimimenetelmän mallin logaritmoidut profiiliuskottavuusfunktioiden kuvaajat lokaatio- ja skaalaparametrille. Kuvaajan maksimi antaa parametrin estimaatin ja minimi estimaatin 99.5% :n luottamusvälin.

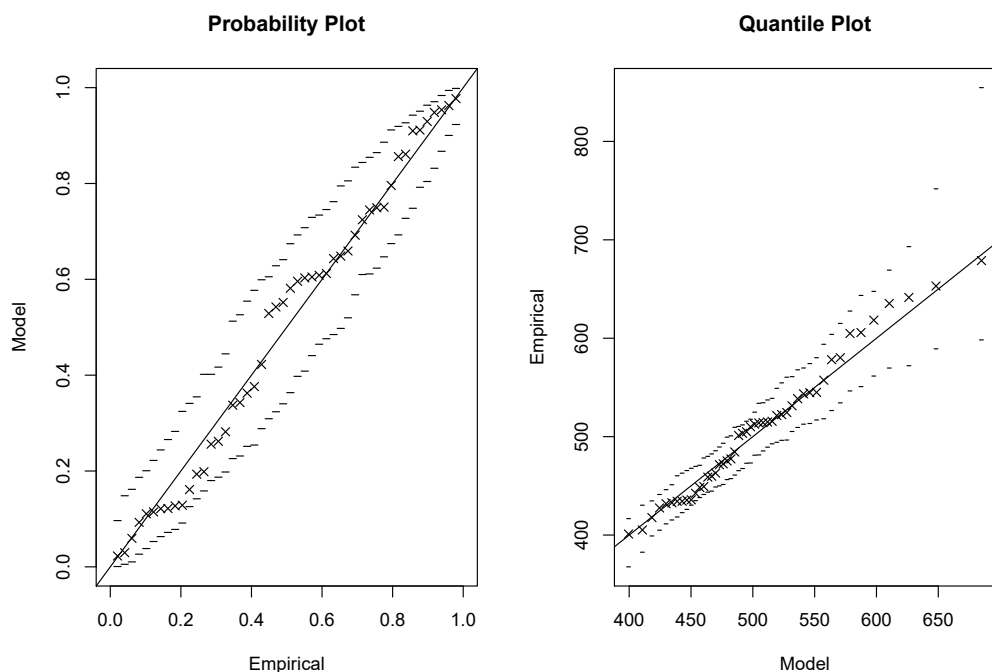
### 4.2.3 Ylitemenetelmä

Sovelletaan aineistoon seuraavaksi ylitemenetelmää. Tätä varten käytetään koko aineistoa, kun blokkimaksimimenetelmässä käytettiin vain vuotuisia maksimeja. Tarkastellaan ensin aineistoa histogrammilla. Kuvasta 12 havaitaan, että havaintojen jakauma on oikealle vino.



Kuva 8: Blokkimaksimimenetelmän mallin logaritmoitu profiiluskottavuusfunktion kuvaaja muotoparametrille. Kuvaajan maksimi antaa muotoparametrin estimaatin ja minimit estimaatin 99.5%:n luottamusvälin. Estimaatin negatiivinen arvo viittaa siihen, että ääriarvot olisivat kuukausittaisten kuolemien tapauksessa Weibull-jakautuneita. Kuvasta havaitaan, että nolla sisältyy luottamusvälien sisälle. Luottamusvälien perusteella ei siis voida sulkea pois vaihtoehtoa, että ääriarvot olisivatkin Gumbel-jakautuneita.



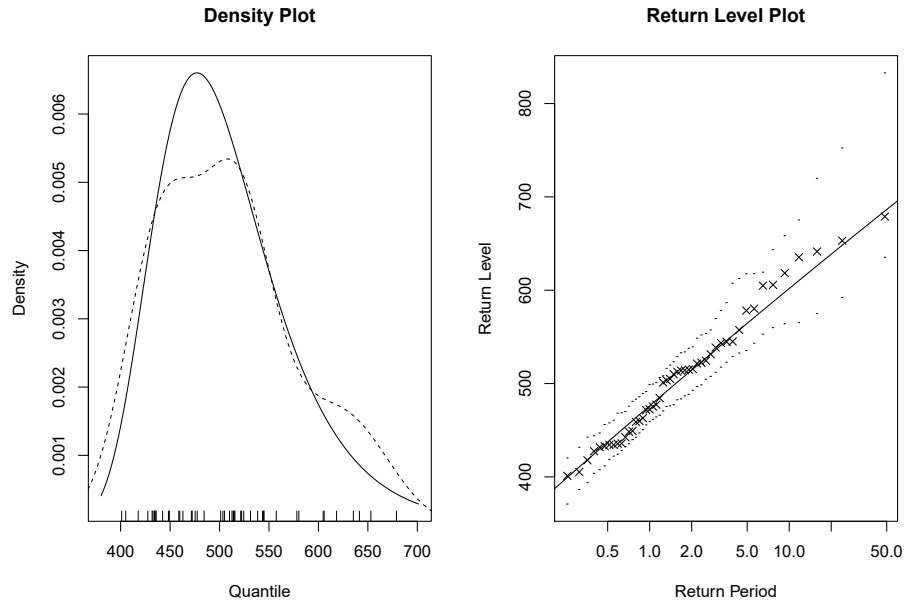


Kuva 9: Vasemmalla blokkimaksimimenetelmän mallin PP-kuvaaja ja QQ-kuvaaja, joissa aineistosta saadut havainnot on merkitty rasteilla ja mallin luottamusvälit viivoilla. Mikäli malli vastaisi aineistoa, sijoittuisivat havainnot molemmissa kuvaajissa lähelle viivaa joka kuvaa arvojen täydellistä vastaavuutta aineistoon. Molemmissa kuvissa kaikki havainnot mukailevat diagonaalia ja ovat luottamusvälien sisällä.

Ensin pitää löytää sopiva kynnyisarvo  $u$ . Osiossa *3.3.2 Yleistetty Pareto-jakauma ja ylitemenetelmä* esitettiin kaksikin tapaa, joiden avulla arvoa  $u$  voidaan etsiä. Kerrotaan vielä lyhyesti, miksi kynnyisarvon valintaan on niin tärkeää kiinnittää huomiota.

Kynnyisarvon valinnassa pitää tasapainoilla estimaattien *harhaisuuden* ja *tarkkuuden* välillä. Jos estimaatti on harhainen se ei vastaa parametriaan, vaikka sen hajonta olisi pientä ja estimaatille laskettavat luottamusvälit olisivat kapeat. Jos taas estimaatti on epätarkka, sen hajonta on suurta ja luottamusvälit ovat leveät, vaikka estimaatti kuvaisi parametria hyvin.

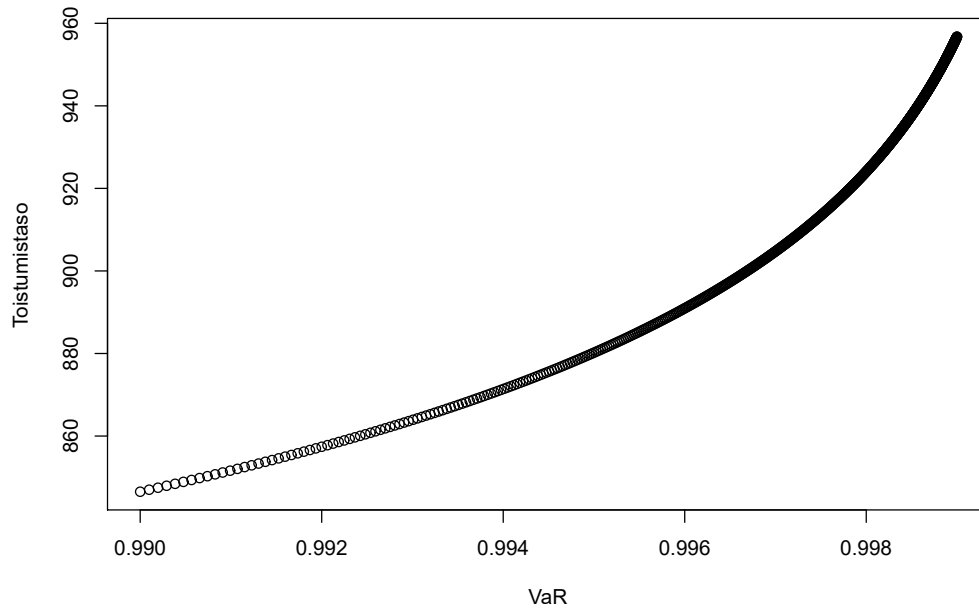
Kynnystaso tulisi valita niin korkeaksi, että sen ylittäviä havaintoja tar-



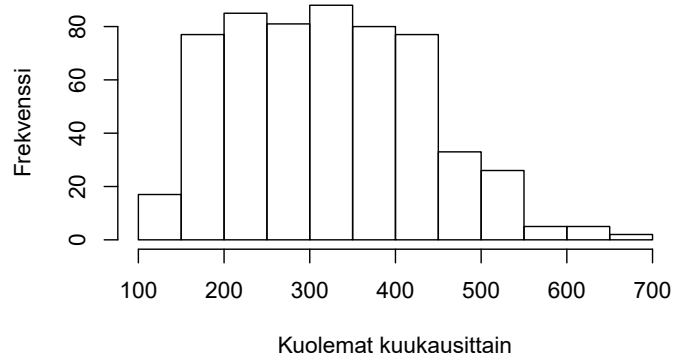
Kuva 10: Kuvassa vasemmalla blokkimaksimimenetelmän antaman mallin tiheysfunktio merkittynä viivalla ja empiirinen tiheysfunktio katkoviivalla. Jos malli on hyvä, kuvaajat ovat lähellä toisiaan. Malli mukailee empiiristä funktiota kohtuullisen hyvin. Kuvassa oikealla toistumistasokuvaaja, jossa viiva kuvaa mallin ja aineiston täyttää vastaavuutta. Aineistosta saadut havainnot on merkitty rasteilla ja mallin luottamusvälit viivoilla. Havainnot mukailevat viivaa kohtalaisesti ja osuvat luottamusvälien sisälle. Kuvan perusteella mallia ei siis voida pitää epäsojivana.

kasteltaessa tarkasteltaisiin todellakin tiheysfunktion häntää. Tällöin havaintojen voidaan olettaa noudattavan yleistettyä Pareto-jakaumaa ja estimaateista saadaan mahdollisimman harhattomia. Toisaalta kynnysarvo halutaan valita mahdollisimman matalaksi, jotta sen yläpuolelle jää tarpeeksi paljon havaintoja. Tällöin estimaateista saadaan tarkempia. Estimaattien luottamusväleistä tulee helposti isoja, jos havaintoja on liian vähän.

Käytetään paketin *evd* [13] funktiota *teplot* sopivan kynnysarvon haaroimiseksi. Funktio estimoi MLE-menetelmää käyttäen muotoparametria  $\mu$  ja muunnettua skaalaparametria  $\sigma^*$  eri kynnysarvoilla  $u$ . Funktiolle annetaan



Kuva 11: Blokkimaksimimenetelmän mallin  $\text{VaR}_{1-\frac{1}{t}}$ . Tässä  $t$  on toistumisjakso vuosina ja  $y$ -akselilla on jaksoa vastaava toistumistaso.



Kuva 12: Histogrammi kuukausittaisista kuolemista. Kuvaajasta havaitaan, että oikea häntä on vasenta pidempi eli havaintojen jakauma on oikealle vino.

syötteenä arvolle  $u$  väli jolla parametreja estimoidaan. Näin kynnsarvoa  $u$  voidaan haarukoida ensin laajemmin ja tarkentaa sitten sopivalta vaikuttavalle välille.

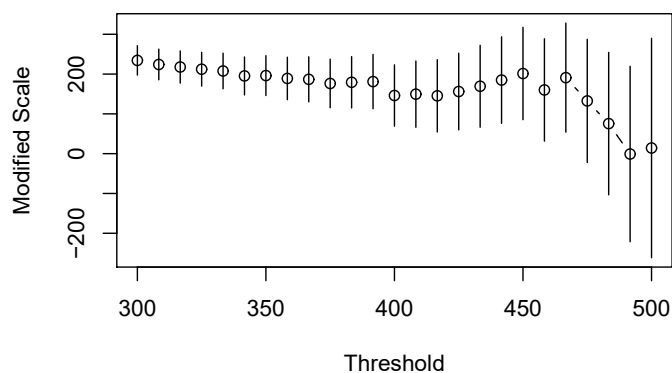
Lähdetään histogrammin ja vuosimaksimien perusteella liikkeelle väliltä  $[300, 500]$ . Kuvissa 13 ja 14 on ympyröillä kuvattu parametrin simuloitua arvoa kyseisellä kynnsarvolla ja viivalla merkitty niiden luottamusväliä. Halutaan valita pienin kynnsarvo  $u$ , josta lähtien molemmat parametrit pysyvät mahdollisimman stabiileina ja niiden luottamusvälit kohtuullisina. Huomataan, että kynnsarvon 450 jälkeen kummankaan parametrin arvo ei pysy enää vakaana ja parametrien luottamusvälit kasvavat suuriksi.

Valitaan kaksi kynnsarvoa, jotta voidaan havainnoida niiden vaikutusta malliin. Suunnilleen kynnsarvosta 340 lähtien molemmat parametrit vaikuttavat stabiileilta ja niiden luottamusvälit kapeilta. Kun katsotaan tarkempaa väliä  $[330, 360]$  kuvissa 16 ja 15 saadaan havainnolle tukea, joten valitaan ensimmäiseksi kynnsarvoksi 340. Valitaan toiseksi kynnsarvoksi 400, sillä kuvista 13 ja 14 huomataan tason lasku tuossa kohtaa, jonka jälkeen parametrit pysyvät taas stabiileina luottamusvälit huomioiden. Olkoon kynnsarvoa  $u = 340$  vastaavaa malli *Malli1* ja kynnstasoa  $u = 400$  vastaava malli *Malli2*.

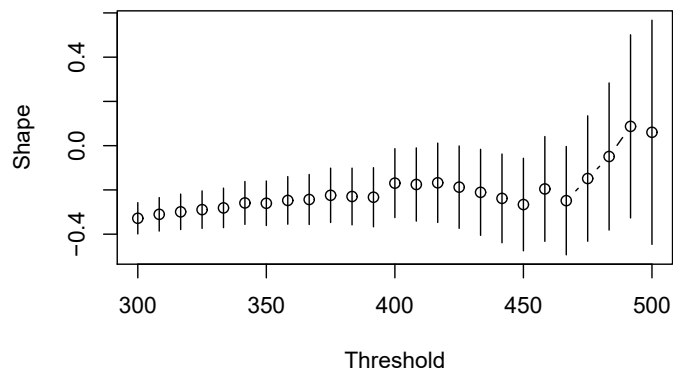
Ensimmäinen kynnsarvo 340 on itse asiassa hyvin matala, sillä kuukausihavainnoista jopa 43% ylittää sen. Toisen kynnsarvon 400 tapauksessa 26% kuukausihavainnoista ylittää kynnyksen. Tämä vaikuttaa edelleen paljolta ottaen huomioon että tarkoitus on mallintaa äärimmäisiä havaintoja. Aineisto on kuitenkin pieni. Siksi tarvitaan isompi osuus kaikista havainnoista, jotta niitä olisi tarpeeksi mallinnusta varten.

Seuraavaksi tehdään itse mallin sovitukset käyttämällä funktiota *fpot*, joka käyttää MLE-menetelmää parametrien estimoimiseksi. Esimaattien luottamusvälejä tarkastellaan funktiolla *profile* ja 99.5%:n luottamusväleillä kuten blokkimaksimimenetelmääkin.

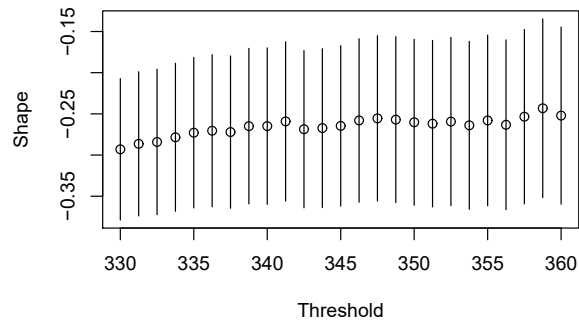
Tarkastellaan ensin kynnsarvoa  $u = 340$  vastaavaa mallia *Malli1*. Tällöin skaalaparametri  $\hat{\sigma} = 108.52$  ja muotoparametri  $\hat{\gamma} = -0.26$ . Kuvassa 17 näkyy estimaattien luottamusvälit ja kuvasta havaitaan, että muotoparametrin luottamusväli ei sisällä arvoa 0. Koska  $\gamma < 0$ , Pareto-jakauma kyn-



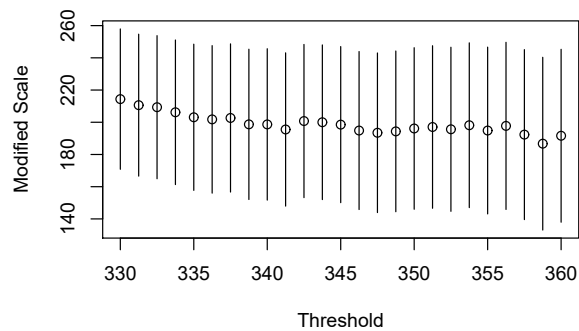
Kuva 13: Muunnettu skaalaparametri  $\sigma^*$  kynnsarvon mukaan välillä 300 – 500. Kuvaan on ympyrällä merkitty parametrin estimaatin arvo ja viivalla sen luottamusväli. Kynnsarvoa valittaessa etsitään sellaista kynnsarvoa, jonka jälkeen muunnettu skaalaparametri pysyisi luottamusväliä huomioiden mahdollisimman vakiona kynnsarvon kasvaessa. Suurilla kynnsarvoilla muunnettu skaalaparametri alkaa heitellä havaintojen vähyden vuoksi ja luottamusväliä kasvavat suuriksi. Tässä tapauksessa muunnettu skaalaparametri pysyy melko tasaisena kynnsarvosta 340 eteenpäin. Toisaalta kynnsarvon 400 kohdalla vaikuttaisi olevan pieni tasomuutos, jonka jälkeen muunnettu skaalaparametri pysyy myös vakaana kunnes alkaa heittelehtiä havaintojen vähyden vuoksi. Siinäkin tapauksessa kynnsarvoa 400 vastaava muunnetun skaalaparametrin arvo sisältyy kyllä luottamusväleihin.



Kuva 14: Muotoparametri kynnsarvon mukaan välillä 300 – 500. Kuvaan on ympyrällä merkitty parametrin estimaatin arvo ja viivalla sen luottamusväli. Kuten muunnetun skaalaparametrin niin myös muotoparametrin estimaatin pitäisi pysyä hyvän kynnsarvon jälkeen luottamusväliä huomioiden mahdollisimman vakiona kynnsarvon kasvaessa. Suurilla kynnsarvoilla havaintoja on vähän, jolloin kynnsarvon kasvaessa muotoparametrin estimaatti alkaa heitellä ja sen luottamusväli kasvat suuriksi. Tässä tapauksessa muotoparametrin estimaatti pysyy melko tasaisena arvosta 340 eteenpäin. Toisaalta kynnsarvon 400 kohdalla vaikuttaisi olevan pieni tasonmuutos, jonka jälkeen estimaatti pysyy myös hetken vakaana. Vaikka estimaatti alkaa heittelehtiä havaintojen vähyyden vuoksi, sisältyy kynnsarvoa 400 vastaava estimaatin arvo kuitenkin luottamusväleihin tällä tarkasteltavalla välillä.

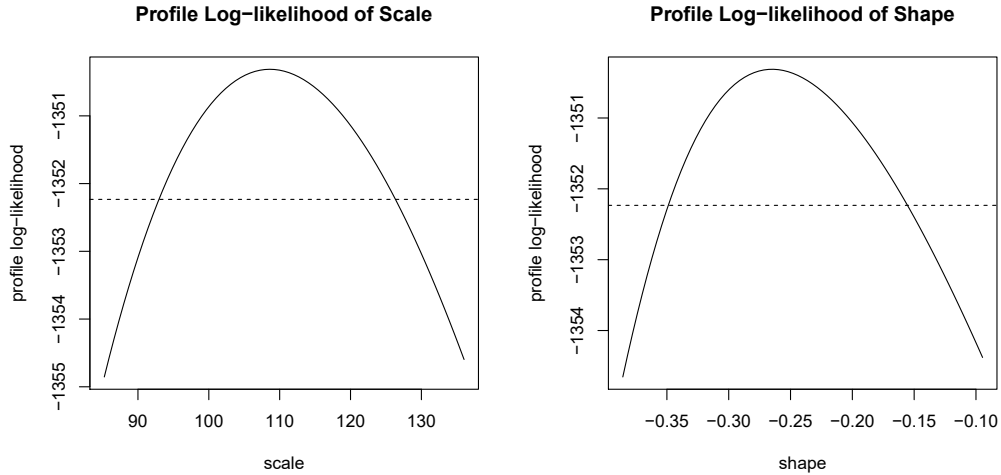


Kuva 15: Tarkastellaan muotoparametria kynnsarvon mukaan tarkemmin, välillä 330–360. Kuvaan on ympyrällä piirretty parametrin estimaatin arvo ja viivalla sen luottamusvälit. Noin kynnsarvosta 340 lähtien muotoparametrin estimaatti vaikuttaa hyvin vakaalta.



Kuva 16: Tarkastellaan muunnettua skaalaparametria  $\sigma^*$  kynnsarvon mukaan tarkemmin, välillä 330 – 360. Kuvaan on ympyrällä piirretty parametrin estimaatin arvo ja viivalla sen luottamusvälit. Kynnsarvon 340 jälkeen muunnettu skaalaparametri pysyy melko tasaisena.

nystason  $u$  ylitteille on lyhythäntäinen ja sillä on äärellinen oikea pääte-  
 piste  $x_F = -\frac{\sigma}{\gamma} = -\frac{108.52}{-0.26} = 416.35$ . Eli tämän mukaan maksimiarvo kuolemille  
 kuukaudessa olisi kokonaisluvuksi pyöristettynä  $x_F + u = 416.35 + 340 \approx 756$ .  
 Arvo vaikuttaa matalalta mallinnettavaa ilmiötä ajatellen.



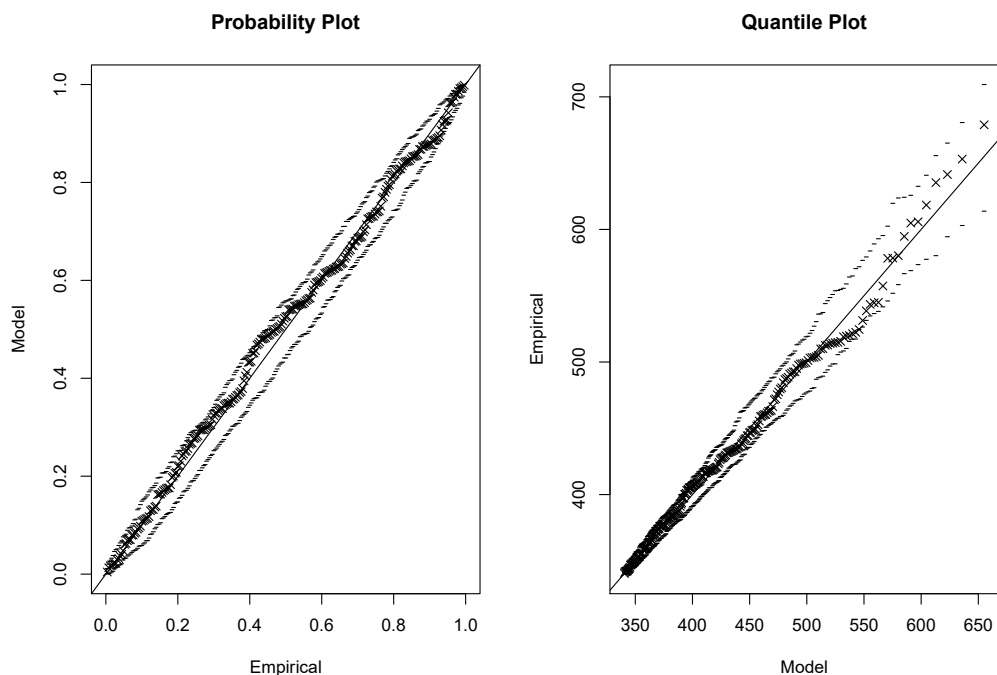
Kuva 17: Ylitemenetelmän malli *Malli1*, jolle  $u = 340$ . Kuvassa mallin loga-  
 ritmoidut profiiliuskottavuusfunktioiden kuvaajat skaala- ja muotoparamet-  
 rille. Kuvaajan maksimi antaa parametrin estimaatin ja kuvaajan minimi  
 estimaatin 99.5% :n luottamusväli. Muotoparametrin estimaatin arvo on  
 negatiivinen, eikä nolla sisälly sen luottamusvälien sisälle. Tämä viittaa sii-  
 hen, että yleistetty Pareto-jakauma kynnsarvon  $u = 340$  ylitteille on lyhyt-  
 häntäinen ja sillä on äärellinen oikea päätepointe.

Tarkastellaan vielä mallin *Malli1* sopivuutta aineistoon piirtämällä funk-  
 tion *f<sub>pot</sub>* tuloste. Saadaan kuvat 18 ja 19. Kuvassa 18 on todennäköisyysku-  
 vaaja ja kvantiilikuvaaja. Jos malli on hyvä, tulisi havaintojen osua diago-  
 naalille molemmissa kuvaajissa. Kuvasta 18 näemme, että ne asettuvat siihen  
 melko hyvin. Kuvaajat eivät siis osoita todisteita mallin sopivuutta vastaan.

Kuvassa 19 on vasemmalla empiirisen ja estimoidun tiheysfunktion ku-  
 vaajat sekä oikealla toistumistason kuvaaja logaritmisella asteikolla vuosina  
 luottamusvälineen. Myös tiheysfunktioiden kuvaajassa malli vaikuttaa mu-  
 kailevan empiiristä tiheysfunktioita kohtuullisen hyvin. Samoin toistumista-

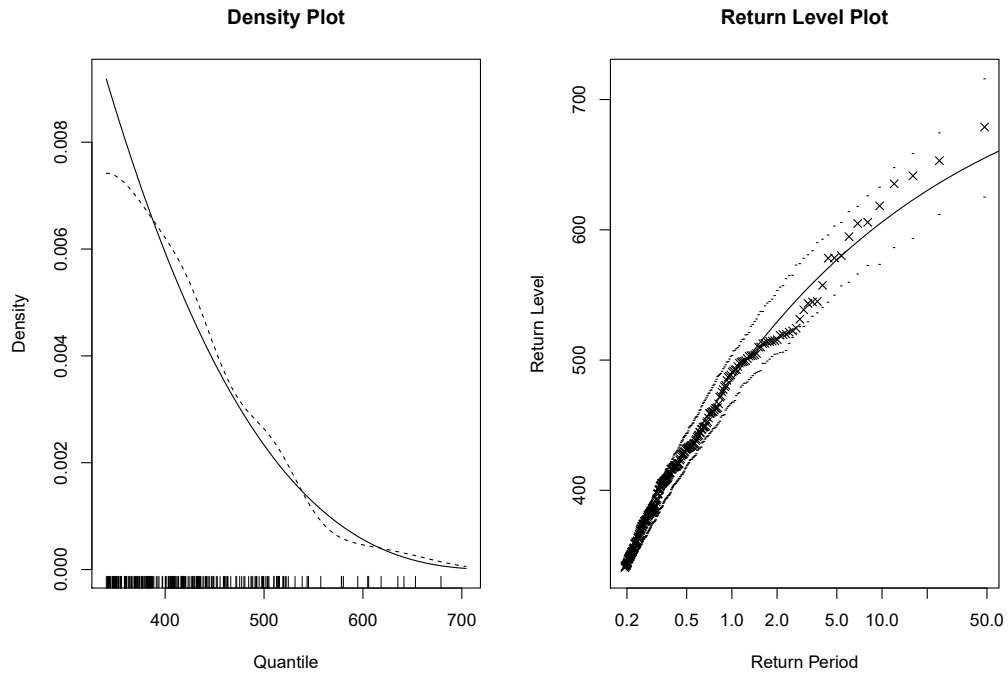


son kuvaajassa havainnot osuvat luottamusvälien sisään ja mukailevat mallin kuvaajaa melko hyvin.

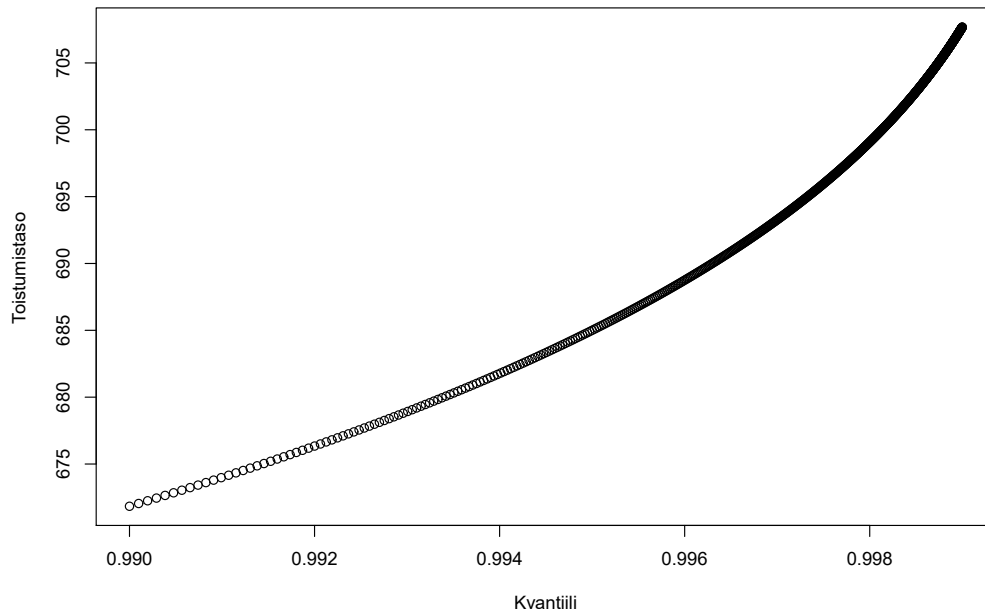


Kuva 18: Ylitemenetelmän *Malli1*, jolle kynnyisarvo  $u = 340$ . PP- ja QQ-kuvaajat, joissa aineistosta saadut havainnot on merkitty rasteilla ja mallin luottamusvälit viivoilla. Mikäli malli vastaisi aineistoa, sijoittuisivat havainnot molemmissa kuvaajissa lähelle diagonaalini viivaa joka kuvaa arvojen täydellistä vastaavuutta aineistoon. Molemmissa kuvissa kaikki havainnot mukailevat diagonaalini. Etenkin QQ-plotin kuvasta on vaikeaa sanoa, ovatko havainnot luottamusvälien sisällä vai eivät. Jos kuvaa suurennettaisiin tarpeeksi niin vaikuttaisi kuitenkin siltä että ne ovat.

Tarkastellaan vielä Value-at-Riskia. Esimerkiksi mallin mukaan  $\text{VaR}_{0.995} = 672$  ja  $\text{VaR}_{0.999} = 708$ . Arvot ovat matalia blokkimaksimimenetelmään verrattuna. Lisäksi jos aineistoa oltaisiin simuloitu suoraan aiemmin kerrotulla kaavalla  $2r \frac{k}{12}$  ilman lisäehtoja, arvo 725 oltaisiin havaittu todennäköisyydellä 0.001. Kuvassa 20 näkyy jakauman kvantiilit eli  $\text{VaR}_{1-\frac{1}{t}}$  laskettuna kaavan (25) mukaisesti missä  $t$  on toistumisjakso vuosina.



Kuva 19: Ylitemenetelmän *Malli1*, jolle kynnyсарvo  $u = 340$ . Kuvassa vasemmalla mallin tiheysfunktio merkittynä viivalla ja empiirinen tiheysfunktio katkoviivalla. Jos malli on hyvä, tiheysfunktiot ovat lähellä toisiaan. Malli mukailee empiiristä funktiota. Kuvassa oikealla toistumistasokuvaaja, jossa viiva kuvaa mallin ja aineiston täyttä vastaavuutta. Aineistosta saadut havainnot on merkitty rasteilla ja mallin luottamusvälit viivoilla. Havainnot mukailevat viivaa kohtalaisesti ja vaikuttavat osuvan myös luottamusvälien sisälle.



Kuva 20: Ylitemenetelmän malli *Malli1*,  $x$ -akselilla  $\text{VaR}_{1-\frac{1}{t}}$  missä  $t$  on toistumisjakso vuosina ja  $y$ -akselilla sitä vastaavat toistumistasot.

Tarkastellaan sitten mallia *Malli2*, jossa  $u = 400$ . Yleistetyn Pareto-jakauman parametrien estimaateiksi saadaan  $\hat{\gamma} = -0.17$  ja  $\hat{\sigma} = 78.11$  ja niiden luottamusvälit näkyvät kuvassa 21. Huomataan että *Malli1*:stä poiketen tällä mallilla 0 kuuluu muotoparametrin  $\hat{\sigma}$  luottamusvälille, kuten blokkimaksimimenetelmän tuottamassakin mallissa. Koska muotoparametrin estimaatti on tälläkin mallilla negatiivinen, on mallin mukaan jakaumalla äärellinen oikea pääte piste  $x_F = -\frac{\sigma}{\gamma} = -\frac{78.11}{-0.17} = 459.47$  jolloin  $x_F + u = 459.47 + 400 \approx 859$ . Maksimiarvo on jo korkeampi kuin *Malli1*:n tapauksessa, mutta vaikuttaa edelleen tarkasteltavaa ilmiötä ajatellen matalalta.

Kuvasta 22 nähdään että sekä PP-kuvaajassa että QQ-kuvaajassa havainnot osuvat luottamusvälien sisään, vaikka vähän heiluvatkin diagonaalin ympärillä. Mallia ei siis voi tämän perusteella pitää epäsojivana.

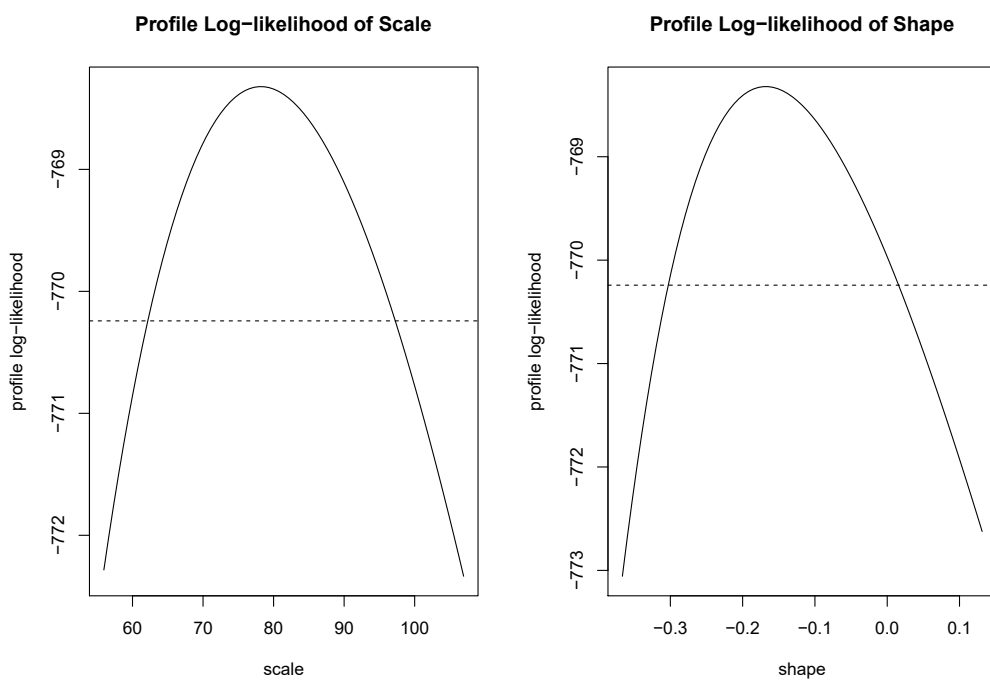
Myös tiheystfunktion ja toistumistason kuvaajista nähdään kuvasta 23, että havainnot mukailevat mallia kohtuullisen hyvin.

Mallilla *Malli2* saadaan  $\text{VaR}_{0.995} = 708$  ja  $\text{VaR}_{0.999} = 745$  eli korkeammat kuin mallilla *Malli1* mutta matalammat kuin blokkimaksimimenetelmällä saadut. Kuvassa 24 on  $\text{VaR}_{1-\frac{1}{t}}$  laskettuna kaavan (25) mukaisesti missä toistumisjakso on vuosia.

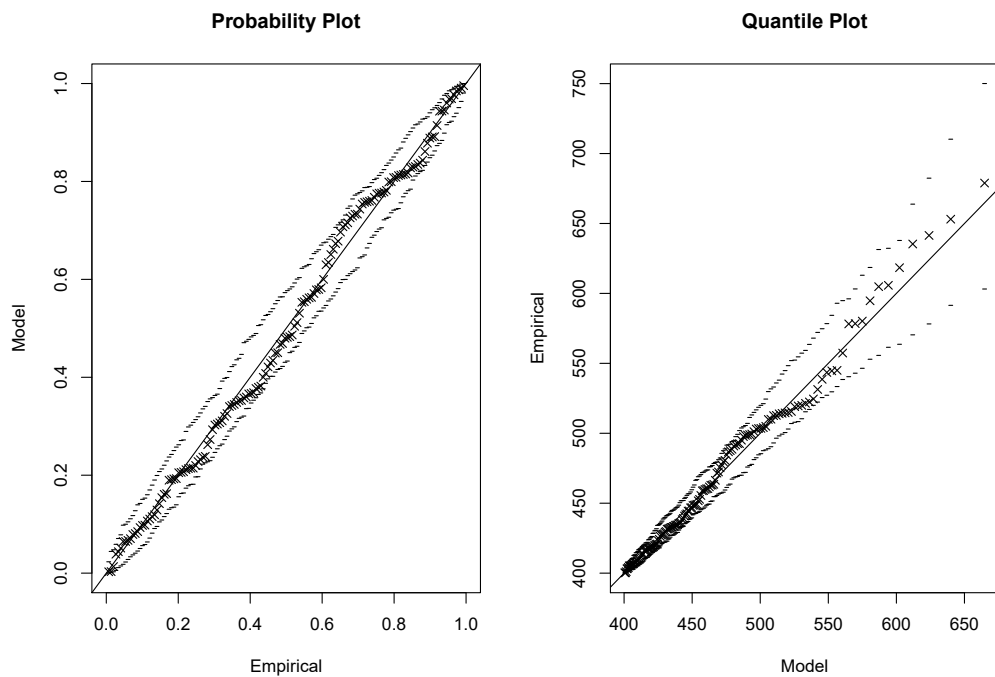
#### 4.2.4 Mallien vertailu ja yhteenveto

Ylitemenetelmän ensimmäiselle mallille valittiin kynnsarvo  $u = 340$ , vaikka todettiin että havainnoista suuri osuus ylittää sen. Mallin sopivuuden tarkasteluun käytettävien kuvaajien 18 ja 19 perusteella malli ei silti vaikuta kelvottomalta. Mallin *Malli2* vastaavista kuvaajista 22 ja 23 nähdään kuitenkin, että toisen mallin tapauksessa havainnot eivät ole yhtä lähellä mallin luottamusvälien reunoja ja vaikuttavat muutenkin mukailevan diagonaalia paremmin kuin ensimmäisessä mallissa. Tällä perusteella voitaisi tässä tapauksessa hylätä sopimattomampi malli *Malli1*.

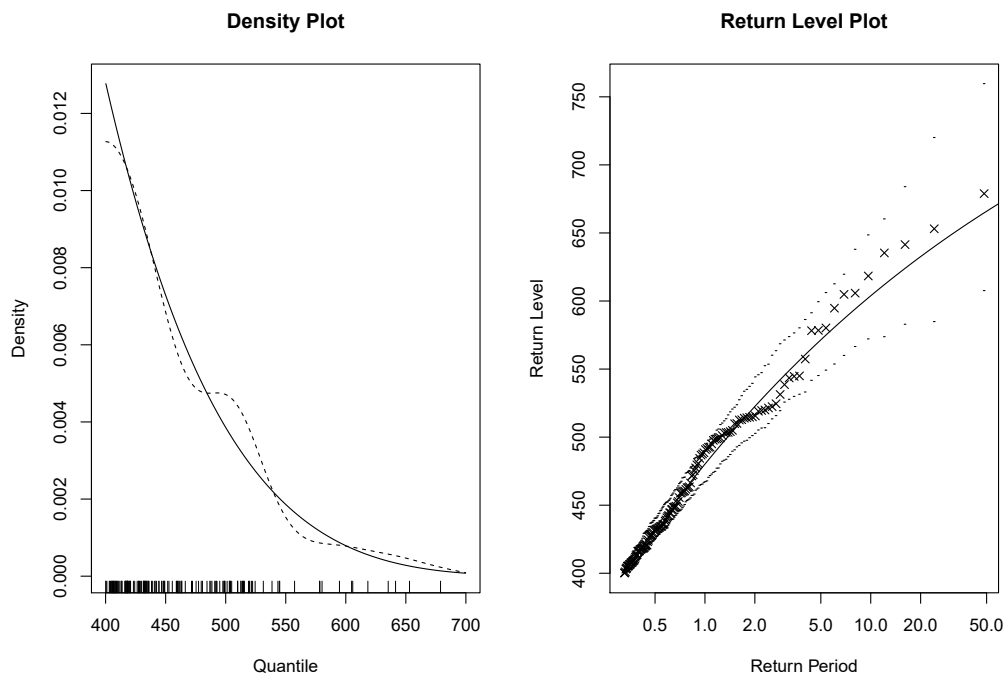
Tiedetään, että kuolemien määrä kuukaudessa saattaisi kasvaa hyvinkin suureksi esimerkiksi sodan takia. Lisäksi yleensä halutaan mieluummin ottaa varman päälle kuin vähätellä riskejä. Siksi korkeampia Value-at-Riskin esti-



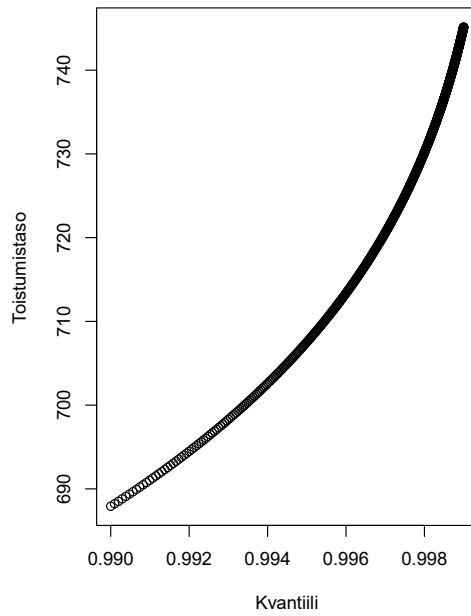
Kuva 21: Ylitemenetelmän malli *Malli2*, jolle  $u = 400$ . Kuvassa mallin logaritmoidut profiiluskottavuusfunktioiden kuvaajat skaala- ja muotoparametreille. Kuvaajan maksimi antaa parametrin estimaatin ja minimi estimaatin 99.5% :n luottamusväli. Muotoparametrin estimaatin arvo on negatiivinen, mutta nolla sisältyy sen luottamusvälien sisälle. Tämä viittaa siihen, että yleistetty Pareto-jakauma kynnysarvon  $u = 400$  ylitteille ei välttämättä ole lyhythäntäinen äärellisellä oikealla päätepisteellä.



Kuva 22: Ylitemenetelmän *Malli2*, jolle kynnyсарvo  $u = 400$ . PP- ja QQ-kuvaajat, joissa aineistosta saadut havainnot on merkitty rasteilla ja mallin luottamusvälit viivoilla. Mikäli malli vastaisi aineistoa, sijoittuisivat havainnot molemmissa kuvaajissa lähelle diagonaalinviivaa joka kuvaa arvojen täydellistä vastaavuutta aineistoon. Molemmissa kuvissa kaikki havainnot mu-kailevat diagonaalia ja havainnot osuvat luottamusvälien sisälle.



Kuva 23: Ylitemenetelmän *Malli2*, jolle kynnyisarvo  $u = 400$ . Kuvassa vasemmalla mallin tiheysfunktio merkittynä viivalla ja empiirinen tiheysfunktio katkoviivalla. Jos malli on hyvä, tiheysfunktiot ovat lähellä toisiaan. Malli mukailee empiiristä funktiota kohtuullisen hyvin. Kuvassa oikealla toistumistasokuvaaja, jossa viiva kuvaa mallin ja aineiston täyttää vastaavuutta. Aineistosta saadut havainnot on merkitty rasteilla ja mallin luottamusvälit viivoilla. Havainnot mukailevat viivaa ja osuvat myös luottamusvälien sisälle.



Kuva 24: Ylitemenetelmän malli *Malli2*,  $x$ -akselilla  $\text{VaR}_{1-\frac{1}{t}}$  missä  $t$  on toistumisjakso vuosina ja  $y$ -akselilla sitä vastaavat toistumistasot.



maatteja antavaa mallia voitaisi pitää parempana. Korkeimmat estimaatit saatiin blokkimaksimimenetelmällä ja *Malli1* antoi matalimmat estimaatit. Tämä tukee päätöstä hylätä *Malli1*, mutta ei yksin riitä pitämään blokkimaksimimenetelmän mallia parhaana.

Vertaillaan vielä blokkimaksimimenetelmän mallia ja ylitemenetelmän mallia *Malli2* keskenään. Niiden muotoparametrien kuvia 8 ja 21 vertaamalla huomataan, että blokkimaksimimenetelmän muotoparametrin luottamusvälit ovat leveämmät kuin ylitemenetelmän mallilla. Blokkimaksimimenetelmän mallin luottamusväleihin siis sisältyy yhtä lyhythäntäinen todennäköisyysfunktio kuin mallissa *Malli2*. Toistumistasokuvaajia 10 ja 23 vertailemalla huomataankin, että blokkimaksimimenetelmän toistumistasojen luottamusvälit ovat leveämmät kuin ylitemenetelmällä. Leveämmät luottamusvälit johtuvat siitä, että blokkimaksimimenetelmä kärsii sekä blokkien sisäisten havaintojen pienestä määrästä (12), että otosten määrän  $m$  vähyydestä (48). Tämä tuo malliin epävarmuutta. Tarkempaa mallia voidaan pitää parempana, olettaen että se on harhaton. Luottamusvälien tarkastelun perusteella ylitemenetelmän *Malli2* saattaisi siis olla tälle aineistolle sopivin malli, vaikka mallilla *Malli1* on itse asiassa pienimmät luottamusvälit. Se ollaan kuitenkin muiden tarkastelujen perusteella todettu epäsovivimmaksi malliksi.

### 4.3 Johtopäätökset

Sekä blokkimaksimimenetelmän että ylitemenetelmän soveltaminen  $R$ -ohjelmointikielellä on käytännössä melko helppoa, sillä sitä varten on tehty useita valmiita paketteja. Koska jokainen näistä paketeista on toteutettu hieman eri tavalla, vaatii niiden funktioihin ja toimintaperiaatteisiin perehtyminen ja itselle sopivimman paketin löytäminen toisaalta melko paljon aikaa. Paketteja käyttäessä saatujen tulosten tulkitseminen voi myös olla vaikeaa, jos ei ole täysin tietoinen siitä, mitä paketin funktiot tekevät ja mitä menetelmiä käyttäen.

Ylitemenetelmää käytettäessä kynnsarvon  $u$  valinta vaikuttaa tuloksiin huomattavan paljon, eikä sen valinta ole yksinkertaista. Lähdemateriaalien

ja kokeilujen perusteella vaikuttaa siltä, että kynnyksarvon valitsemiseen käytettävistä kuvaajista löytyy käytännössä harvoin selkeää kohtaa, josta optimaalisen kynnyksarvon voisi valita. Käytännössä parhaan kynnyksarvon löytämiseksi on paras haarukoida sopivaa arvoa ensin mahdollisimman hyvin ja tehdä sitten mallinnus muutamalla eri kynnyksarvolla ja vertailla saatuja tuloksia keskenään.

Käytetylle aineistolle sekä blokkimaksimimenetelmällä että ylitemenetelmällä luotujen mallien muotoparametrien luottamusvälit sisältävät vaihtoehdon, että mallin jakaumalla olisi äärellinen oikea päätepiste. Toiselle ylitemenetelmän malleista muotoparametrin luottamusvälien sisään ei osu edes muuta vaihtoehtoa. Kuolemien määrä on todellisuudessaakin rajoitettu elävien ihmisten lukumäärällä. Koska periaatteessa kaikki voisivat kuolla yhden kuukauden aikana, on todellinen yläraja kuitenkin paljon korkeampi kuin mallien antama. Toki tällaisen tapahtuman todennäköisyys on käytännössä nolla.

Kuolemien määrän ennustaminen suoraan onkin sikäli haastavaa, että se on voimakkaasti riippuvainen muista satunnaisista ilmiöistä kuten luonnonkatastrofeista. Lisäksi pitkällä aikavälillä tarkasteltuna kuolevuus ei välttämättä ole stationaarinen prosessi.

Tilastokeskuksen tilastossa on melko vähän vaihtelua vuosien välillä. Se kuvaa paremmin normaalia kuolevuutta, sillä ainoa tavanomaista enemmän kuolemia aiheuttanut tapahtuma joka osuu tarkasteluvälille on 90-luvun lama. Vuosittaiset kuolemat voivat olla paljon normaalia suurempia esimerkiksi vallitsevan Covid-19 epidemian aikana. Aineiston vähäisestä vaihtelusta johtuen mallit tuottavat myös estimaateille  $\text{VaR}_{0.999}$  ja  $\text{VaR}_{0.995}$  näin ajateltuna melko alhaisilta tuntuvat arvot. Vaikka mallit olivat aineistoon melko sopivia, ne tuntuvat kuitenkin aliarvioivan riskiä. Leppisaari [9] muistuttaakin teoksessaan, että aineiston ulkopuolelle ekstrapolointi tuottaa aina paljon epävarmuutta ja erityisesti mitä pidemmälle ekstrapoloidaan. Siksi hän toteaaakin, että vahvojen johtopäätösten kanssa tulee olla hyvin varovainen.

## Lähteet

- [1] Coles, S. G. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- [2] de Haan L. ja Ferreira A.F. 2006. *Extreme Value Theory An Introduction*. Springer-Verlag New York.
- [3] Dowd K. ja Blake D. 2006. *After VaR: The Theory, Estimation, and Insurance Applications of Quantile-Based Risk Measures*. The University of Nottingham. United Kingdom.
- [4] Euroopan komissio. 2015. *Solvency II Overview – Frequently asked questions*. [https://ec.europa.eu/commission/presscorner/detail/en/MEMO\\_15\\_3120](https://ec.europa.eu/commission/presscorner/detail/en/MEMO_15_3120). Luettu 16.5.2020.
- [5] Finanssivalvonta. 2018. *Solvenssi II*. <http://www.finanssivalvonta.fi/fi/Saantely/Saantelyhankkeet/Solvenssi/Pages/Default.aspx>. Sivü päivitetty 20.4.2018. Luettu 13.7.2018.
- [6] Finanssivalvonta, Tiainen S. 2011. *Suomalaisten vakuutusyhtiöiden Solvenssi II-vaikuttavuustutkimus*. 15.3.2011. <http://www.finanssivalvonta.fi/fi/Tiedotteet/Lehdistotiedotteet/Documents/QIS5.pdf>. Luettu 13.7.2018
- [7] Group of 30. 1993. *Derivatives: Practice and Principles*. Group of Thirty. Washington.
- [8] Holton G.A. 2014. *Value-at-Risk: Theory and practice*. Holton G.A. Toinen painos. Luku 11.
- [9] Leppisaari M. 2013. *Äärimmäisten ilmiöiden mallintamisesta*. SHV-työ.
- [10] Moorad C. 2013. *An introduction to Value-at-Risk*. John Wiley & Sons Ltd. United Kingdom.
- [11] Morgan J.P. *RiskMetrics - Technical Document*. Morgan Guaranty Trust Company. New York.

- [12] Owen A.B. 2013. *Monte Carlo Theory, Methods and Examples*. Owen A.B. Luvut 1-2.
- [13] Stephenson A. G. 2002. *evd: Extreme Value Distributions*. R News, 2(2):31-32, June 2002. URL: <https://CRAN.R-project.org/doc/Rnews/>, <https://mran.microsoft.com/snapshot/2015-05-02/web/packages/evd/evd.pdf>
- [14] Suomen virallinen tilasto (SVT). Kuolemansyyt [verkkojulkaisu]. ISSN=1799-5051. Helsinki: Tilastokeskus [viitattu: 27.11.2018]. Saantitapa: <http://www.stat.fi/til/ksyyt/index.html>

## A Käytetty R-ohjelmointikielen koodi

```
#yleistetyn ääriarvojakauman todennäköisyysfunktiot
library(evd)
plot(seq(0,6,0.01),dgev(seq(0,6,0.01),loc=0, scale=0.38, shape=0),
type = "l", ylab='', xlab = '', ylim = c(0,1))

par(new=TRUE)
plot(seq(0,6,0.01),dgev(seq(0,6,0.01),loc=0, scale=0.38, shape=-0.25),
type = "l", col = 'red', ylab='', xlab = '', ylim = c(0,1))

par(new=TRUE)
plot(seq(0,6,0.01),dgev(seq(0,6,0.01),loc=0, scale=0.38, shape=1) ,
type = "l", col = 'blue', ylab='', xlab = '', ylim = c(0,1))

legend("topright",
      legend = c("Gamma = 0", "Gamma < 0", "Gamma > 0"),
      col = c('black', 'red', 'blue'),
      pch = c(20,20,20),
      bty = "n",
      pt.cex = 2,
      cex = 1,
      text.col = "black",
      horiz = F ,
      inset = c(0.1, 0.1))

#yleistetyn ääriarvojakauman kertymäfunktiot
plot(seq(0,6,0.001),pgev(seq(0,6,0.001),loc=1, scale=0.5, shape=0),
type = "l", ylab='', xlab = '', ylim = c(0,1))

par(new=TRUE)
plot(seq(0,6,0.001),pgev(seq(0,6,0.001),loc=1, scale=0.5, shape=-0.25),
type = "l",col = 'red', ylab='', xlab = '', ylim = c(0,1))
```

```

par(new=TRUE)
plot(seq(0,6,0.001),pgev(seq(0,6,0.001),loc=1, scale=0.5, shape=1) ,
type = "l",col = 'blue', ylab='', xlab = '', ylim = c(0,1))

legend("bottomright",
      legend = c("Gamma = 0", "Gamma < 0", "Gamma > 0"),
      col = c('black', 'red', 'blue'),
      pch = c(20,20,20),
      bty = "n",
      pt.cex = 2,
      cex = 1,
      text.col = "black",
      horiz = F ,
      inset = c(0.1, 0.1))

#yleistetyn paretojakauman tiheysfunktio
plot(seq(0.01,6,0.01),dgpd(seq(0.01,6,0.01),loc=0, scale=1, shape=0),
type = "l",ylab='', xlab = '', xlim=c(0,6),ylim = c(0,1.1))

par(new=TRUE)
plot(seq(0.01,6,0.01),dgpd(seq(0.01,6,0.01),loc=0, scale=1, shape=-0.25),
type = "l",col = 'red', ylab='', xlab = '', xlim=c(0,6),ylim = c(0,1.1))

par(new=TRUE)
plot(seq(0.01,6,0.01),dgpd(seq(0.01,6,0.01),loc=0, scale=1, shape=1) ,
type = "l",col = 'blue', ylab='', xlab = '',xlim=c(0,6),ylim = c(0,1.1))

legend("topright",
      legend = c("Gamma = 0", "Gamma < 0", "Gamma > 0"),
      col = c('black', 'red', 'blue'),
      pch = c(20,20,20),
      bty = "n",

```

```

    pt.cex = 2,
    cex = 1,
    text.col = "black",
    horiz = F ,
    inset = c(0.1, 0.1))

#yleistetyn paretojakauman kertymäfunktio
plot(seq(1,6,0.001),pgpd(seq(1,6,0.001),loc=1, scale=1, shape=0),
type = "l", ylab='', xlab = '', ylim = c(0,1))

par(new=TRUE)
plot(seq(1,6,0.001),pgpd(seq(1,6,0.001),loc=1, scale=1, shape=-0.25),
type = "l", col = 'red', ylab='', xlab = '', ylim = c(0,1))

par(new=TRUE)
plot(seq(1,6,0.001),pgpd(seq(1,6,0.001),loc=1, scale=1, shape=1) ,
type = "l", col = 'blue', ylab='', xlab = '', ylim = c(0,1))

legend("bottomright",
      legend = c("Gamma = 0", "Gamma < 0", "Gamma > 0"),
      col = c('black', 'red', 'blue'),
      pch = c(20,20,20),
      bty = "n",
      pt.cex = 2,
      cex = 1,
      text.col = "black",
      horiz = F ,
      inset = c(0.1, 0.1))

#####
#ladataan aineisto
mydata <- read.csv("C:/Users/Name/Documents/Gradu_kuolinsyyt.csv",
sep = ",")

```

```

View(mydata) #katsotaan miltä se näyttää

#ladataan paketteja muotoiluja varten
library(plyr)
#library(dplyr)
library(readr)
library(reshape2)
library(stringr)

#tehdään aineisto data frameksi
kuolemat <- data.frame(mydata[3:50])
kuolemat<-data.frame(kuolemat[1,])
kuolemat2<-as.list((t(kuolemat)))
t(kuolemat2)/12

#luodaan data.frame kkuolemat keinotekoisen kuukausijaottelun luomiseksi.
#luodaan kuukausivektori, johon lisätään satunnaisia osuuksia kuolemista,
#kunnes vuosikiintiö on täynnä
kkuolemat<-data.frame(matrix(c(seq(-2,-1,length.out=576)), nrow = 48,
ncol=12))

vuodet<-c()
for (i in 1:length(kuolemat)){
  kuukausi<-c()
  j<-as.numeric(kuolemat[i])
  print(j) #testi
#niin kauan kuin kuukausissa on yhteensä alle 0.95*kuolemantapaukset
  while(as.numeric(sum(kuukausi[]))<(0.95*j) | length(kuukausi)<12){

#jos vektori on alle 12kk ja summa ei ole täynnä
  if(length(kuukausi)<12 & as.numeric(sum(kuukausi[]))<0.95*j)){

```



```

    k<-0
    while (k<(j/22)){
      k<-(runif(1)*2*(j/12))
      print(k)
    }
    kuukausi<-c(kuukausi, k) #random[0,1] *1.5* vuoden kuolemat
#12kuukaudelle
  }
  if(length(kuukausi)==12 & as.numeric(sum(kuukausi[])<0.95*j)){
#jos kuukaudet on täynnä, mutta kuolleiden summa ei, lisätään joulukuuhun
#loput
    kuukausi[12]<-as.numeric(kuukausi[12])+(j-as.numeric(sum(kuukausi[])))
  }
#jos summa on täynnä mutta vuosi ei
  if(length(kuukausi)<12 & as.numeric(sum(kuukausi[])>=0.95*j)){
    kuukausi<-kuukausi*0.95#pienennetään kuukausia
  }
}
  kkuolemat[i,]<-(kuukausi)
}
kuukausi
sum(kuukausi)

View(kkuolemat) #katsotaan luotua dataframea
#laitetaan jokaisen vuoden maksimi viimeiseksi sarakkeeksi
kkuolemat$max<-apply(kkuolemat, 1, max)
#tallennetaan aineisto tiedostoksi
write.csv(kkuolemat, file = "GraduSimuAineisto.csv",row.names=FALSE)

#####

kkuolemat <- read.csv("C:/Users/Name/Documents/GraduSimuAineisto.csv",
sep = ",")

```

```

maxk<-kkuolemat$max
imaxk<-as.integer(maxk)

#luodaan dataframe ilman maxeja aikasarjan piirtämistä varten.
drops <- c("max")
#luodaan dataframe ilman maxeja
kkuolemat_maxiton <- kkuolemat[ , !(names(kkuolemat) %in% drops)]
#tehdään transpoosi jotta saadaan vuodet sarakkeisiin piirtämistä
#ja jatkokäsittelyä varten
kkuolemat_maxiton <- as.data.frame(t(kkuolemat_maxiton))
colnames(kkuolemat_maxiton)<-factor(c(1969:2016)) #laitetaan vuodet
# sarakkeiden nimiksi

head(kkuolemat_maxiton)

#etsitaan maksimien indeksit piiramista varten
indeksit<-c()
for(i in 1:48){
  indeksit <- c(indeksit,which(c(ts(kkuolemat_maxiton))
==c(kkuolemat$max)[i]))
}
indeksit
#Piiirretään aikasarja kaikista kuukausihavainnoista

plot(ts(as.integer(c(ts(kkuolemat_maxiton)))),
main="Kuolemat kuukausittain 1969-2016",
ylab='Kuolemat', xlab = 'Aika',xlim=c(1,576),ylim=c(100,700),xaxt="n")

par(new=TRUE)
plot(indeksit, c(kkuolemat$max), col = 'red', ylab='',

```

```

xlab = '',xlim=c(1,576), ylim=c(100,700) ,xaxt="n")
axis(1, at=c(1,288,564), c(1969,1993,2016))

#Piiirretään ylitemenetelmä

#havainnot jotka ovat yli kynnysarvon u
kkuolemat_yliu<-kkuolemat_maxiton[(kkuolemat_maxiton>400)==TRUE]
indeksit2<-c()
for(i in 1:length(kkuolemat_yliu)){
  indeksit2 <- c(indeksit2,which(c(ts(kkuolemat_maxiton))
==c(kkuolemat_yliu)[i]))
}
plot(ts(as.integer(c(ts(kkuolemat_maxiton))))),
main="Kuolemat kuukausittain 1969-2016", ylab='Kuolemat',
xlab = 'Aika',xlim=c(1,576),ylim=c(100,700),xaxt="n")

par(new=TRUE)
plot(indeksit2, c(kkuolemat_yliu), col = 'red', ylab='', xlab = '',
xlim=c(1,576), ylim=c(100,700) ,xaxt="n")

par(new=TRUE)
plot(c(1:576),c(rep(400,576, length_out = 577)), xlim=c(1,576),
ylim=c(100,700), col = 'red', ylab='', xlab = '' ,xaxt="n",
type='l')

axis(1, at=c(1,288,564), c(1969,1993,2016))
length(c(1969:2016))*12

##Analyysikoodi

#blokkimaksimimenetelmä

```

```

par(mfrow = c(1,1))
estimaatit2 <- fgev(kkuolemat$max) #funktio mallin luomiseksi
plot(estimaatit2)
#profile(estimaatit2)
conf1<-profile(estimaatit2, conf = 0.995)
plot(conf1)
confint(profile(estimaatit2, conf = 0.995), level = 0.995)
#conf2 <- profile2d(estimaatit2, conf1, which = c("scale", "shape"))
#plot(conf2)
mu2 = 476.15769
sigma2 = 55.72518
xi2 = -0.01843
years = 100:1000
plot(1-1/years, mu2+(sigma2/xi2)*((-log(1-1/(years*12))))^(-xi2)-1),
     xlab = "VaR", ylab = "Toistumistaso")
var = 1-1/(1000*12)
mu2+(sigma2/xi2)*((-log(var))^(-xi2)-1)
var = 1-5/(1000*12)
mu2+(sigma2/xi2)*((-log(var))^(-xi2)-1)

#kokeillaan millainen malli saadaan, jos alkuarvoksi annetaan gev.fitin
#antama optimipiste.
#Tulos: saadaan gev.fitin optimipiste. Malli on hyvin lähellä estimaatti2:ta
#malli on siis oleellisesti sama kuin estimaatit2
estimaatit3 <- fgev(kkuolemat$max, start=list(loc = 476.197, scale = 55.738,
shape = -0.01866))
conf1_3<-profile(estimaatit3, conf = 0.995)
plot(conf1_3)
confint(profile(estimaatit3, conf = 0.995), level = 0.995)
conf2_3 <- profile2d(estimaatit3, conf1_3, which = c("scale", "shape"))
plot(conf2_3)
citation("evd")

```

```

#ylitemenetelmä
library(evd)

#tutkitaan erilaisia menetelmiä ylitteen löytämiseksi.
#Ensin gpd.fitrage, etsitään (ensimmäistä) kohtaa jossa parametrit
#olisivat stabiileja
kkuolemat_vec <-as.vector(unlist(kkuolemat_maxiton)) #tehdään datasta
#vektori analyyseja varten
gpd.fitrage(kkuolemat_vec,300,500,100) #haarukoidaan stabiilia kohtaa
gpd.fitrage(kkuolemat_vec,300,400,100)#300 ja 400 väliltä tarkemmin
gpd.fitrage(kkuolemat_vec,340,360,50)#340 ja 360 voisi olla
gpd.fitrage(kkuolemat_vec,340,345,50)#340 vaikuttaa jo hyvin stabiililta
gpd.fitrage(kkuolemat_vec,350,550,100)#340 vaikuttaa jo hyvin stabiililta

par(mfrow = c(1,1)) #vaihta 1,2 jos haluat kuvat vierekkäin
tcplot(kkuolemat_vec, tlim = c(300,500)) #kuva
tcplot(kkuolemat_vec, tlim = c(330,360)) #kuva

#Mallit
Model1<-fpot(kkuolemat_vec, 340, model= "gpd", npp = 12)
Model2 <- fpot(kkuolemat_vec, 400, model= "gpd", npp = 12)
Model3 <- fpot(kkuolemat_vec, 450, model= "gpd", npp = 12) #testimalli
#mean residual life plots.
#Näistä voi katsoa todennäköisyyttä että kynnyсарvo on ylitetty,
#jos kyseinen arvo havaitaan
mrlplot(kkuolemat_vec, tlim =c(300,450))
#kuvaajan lineaarisuus viittaa u>u0 viittaa ylitteiden noudattavan
#gp-jakaumaa
#pscale=TRUE näyttää kynnyksen ylittymisen todennäköisyyden
#kynnyksen sijaan
mrlplot(kkuolemat_vec,pscale=TRUE)

```

```

par(mfrow = c(1,2))

plot(Model1)#kuva
plot(Model2)#kuva
plot(Model3)
Model1Plot <- profile(Model1, conf=0.995) #kuva
Model2Plot <- profile(Model2, conf=0.995) #kuva
plot(Model1Plot) #kuva, vaikuttaa siltä että muotoparametri ei sisällä 0:aa.
plot(Model2Plot) #kuva, muotparametrin luottamusväli sisältää 0:n

#maksimiarvot mallien mukaan
-Model1$estimate["scale"]/Model1$estimate["shape"]+340
-Model2$estimate["scale"]/Model2$estimate["shape"]+400

toistumistaso <- function(data, u, beta, ksi, n, N){
  if(ksi==0){
    return((u+beta*log(n*N*length(data[(data>u)==TRUE])/length(data))) )
  }
  else{
    return(u+(beta/ksi)*((n*N*(length(data[(data>u)==TRUE])
/length(data)))^ksi-1))
  }
}

toistumistaso(kkuolemat_vec, 340, 108.52, -0.2644,12,50)
  59*length(kkuolemat_vec[(kkuolemat_vec>678)==TRUE])
/length(kkuolemat_vec)

#toistumistason kuvaajat
years = 100:1000
plot(1-1/years, toistumistaso(kkuolemat_vec, 340, 108.52, -0.2644,

```

```

12,years),
xlab = "Kvantiili", ylab = "Toistumistaso" )

plot(1-1/years, toistumistaso(kkuolemat_vec, 400, 78.1136, -0.1673,
12,years),
xlab = "Kvantiili", ylab = "Toistumistaso" )

#tästä otettu lukuja käyttöön
toistumistaso(kkuolemat_vec, 340, 108.52, -0.2644,12,1000)
#tästä otettu lukuja käyttöön
toistumistaso(kkuolemat_vec, 400, 78.1136, -0.1673,12,1000)

#mallin tiheysfunktio
plot(1:400,1-dgpd(1:400, loc=0, scale=108.5200, shape= -0.2644))
#mallin tiheysfunktio
plot(1:400,qgpd(dgpd(1:400, loc=0, scale=78.1136, shape= -0.1673),
loc=0, scale=78.1136, shape= -0.1673))

#mallin todennäköisyysfunktio
1-plot(1:400,pgpd(1:400, loc=0, scale=108.5200, shape= -0.2644))
plot(300:700, (108.52-0.2644*(300:700))/1.2644)
plot(1:400,pgpd(1:400, loc=0, scale=(108.5200+0.2644*340), shape= -0.2644))

#histogrammi frekvenssistä
hist(as.vector(unlist(kkuolemat_maxiton)),main= " ",
xlab="Kuolemat kuukausittain", ylab = "Frekvenssi")

#vertailua simuloidun ja alkuperäisen aineiston välillä
#luodaan ensin vektori jossa summataan kuukausittainen data vuositasolle
vkuolemat<-c()
sum(kkuolemat_maxiton[2,])
for (i in 1:48 ){

```

```

vkuolemat <- c(vkuolemat,sum(kkuolemat_maxiton[i,]))
}
#katsotaan miten paljon simuloitu aineisto maksimissaan heittää
#alkuperäisestä
max(((vkuolemat-kuolemat)/kuolemat))
#katsotaan miten paljon simuloitu aineisto minimissään heittää
#alkuperäisestä
min(((vkuolemat-kuolemat)/kuolemat))

max(kuolemat/12*2) #kuukausittainen maksimi
#noin todennäköisyys että saataisiin kuolematilaston
#potentiaalinen kuukausittainen maksimi ilman kaatoluokkaa, 725.16
1-dbinom(0, size=12,prob=0.001)

#Kuvaaja alkuperäisestä aineistosta
plot(c(1969:2016),c(kuolemat), xlab= 'Vuodet', ylab = 'Kuolemat') #käytetty.
#Alkuperäinen aineisto
boxplot(t(kuolemat))

#ylitemenetelmän kynnsarvojen ylittävien havaintojen määrä käytetyissä
#malleissa
kkuolemat_maxiton
length(kkuolemat_maxiton[(kkuolemat_maxiton>400)==TRUE])
length(kkuolemat_maxiton[(kkuolemat_maxiton>340)==TRUE])

```