

Tilastotieteen pro gradu -tutkielma

Elinaikojen mallinnus kvantiilisekoitusten avulla

Joonas Itkonen
Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
15. syyskuuta 2014

JYVÄSKYLÄN YLIOPISTO

Matematiikan ja tilastotieteen laitos

Itkonen, Joonas : Elinaikojen mallinnus kvantiilisekoitusten avulla

Tilastotieteen pro gradu -tutkielma, 46 sivua + liitteet 7 sivua,

15. syyskuuta 2014

Tiivistelmä

Polynomiset kvantiilisekoitukset on yhden muuttujan parametrinen malliperhe, jossa mallinnetaan jakauman kvantiilifunktio lineaarikombinaationa ns. pohjajakau- man kvantiilifunktiosta ja toisen asteen polynomista. Tämä generoi tiettyjen eh- tojen täyttyessä todennäköisyysjakauman. Tällaisen mallin idea on lisätä pohjaja- kaumaan esimerkiksi vinous- ja huipukkuuskomponentteja kvantiilifunktioon lisätyn polynomin muodossa.

Kvantiilisekoitusmallia voi soveltaa elinaikojen mallinnukseen mallintamalla elinai- kojen logaritmia tällaisilla kvantiilisekoituksilla. Tämä antaa eksponenttimuunnok- sella mallin alkuperäiselle elinaikajakaumalle. Logaritmiskaalan pohjafunktioksi voi- daan valita esimerkiksi normaalijakauma, logaritminen eksponenttijakauma tai lo- garitminen Gamma(1, 2)-jakauma, joiden generoimat elinaikaskaalan mallit saavat nimet lognormaali-polynominen, eksponentti-polynominen ja gamma-polynominen kvantiilisekoitus.

Huomataan, että näiden mallien parametrit voidaan estimoida esimerkiksi niin sa- nottujen L-momenttien avulla, suurimman uskottavuuden menetelmän tai kvan- tiilifunktioon sovelletun pienimmän neliösumman menetelmän avulla. Simuloimal- la jakaumia ja vertaamalla niitä Weibull-jakaumaan nähdään, että lognormaali- polynominen malli voi olla heikohko elinaikojen mallinnuksessa, mutta kaksi viimeis- tä toimivat paremmin, ja ne ovat varsin joustavia malleja, joilla voi mallintaa hy- vin erilaisia elinaika-aineistoja. Erityisesti havaitaan, että eksponentti-polynominen malli yleistää Weibull-jakauman.

Soveltamalla näitä malleja suomalaisten käytössä olevien matkapuhelinten ikära- kennetta kuvaavaan aineistoon nähdään muun muassa, että vanhemmilla ihmisillä käytössä olevat puhelimet ovat keskimäärin vanhempia.

Avainsanoja: elinaika, elinaika-analyysi, kvantiilifunktio, kvantiilisekoitus, lognor- maali, eksponentti, parametrinen, L-momentit, Weibull-jakauma, matkapuhelin

Sisältö

1	Johdanto	1
2	Kvantiilisekoitukset	2
2.1	Yleinen kvantiilisekoitus	2
2.2	Polynominen kvantiilisekoitus	2
3	Kvantiilisekoitusten parametrien estimointi	6
3.1	Suurimman uskottavuuden menetelmä	6
3.2	L-momenttimenetelmä	7
3.2.1	L-momentit	7
3.2.2	Parametrien estimointi L-momenttien avulla	8
3.3	Kvantiilifunktion pienimmän neliösumman menetelmä	9
4	Polynominen kvantiilisekoitusmalli elinajoille	10
4.1	Lyhyesti elinajoista	10
4.2	Kvantiilisekoitusmallin soveltaminen elinaikoihin	10
4.3	Lognormaali-polynominen kvantiilisekoitus	11
4.3.1	Simulointiesimerkki	12
4.4	EkspONENTTI-polynominen kvantiilisekoitus	13
4.4.1	Yhteys Weibull-jakaumaan	14
4.4.2	Simulointiesimerkki	16
4.5	Gamma-polynominen kvantiilisekoitus	17
4.5.1	Simulointiesimerkki	18
5	Kvantiilisekoitusten sovittumistarkasteluja	20
5.1	Weibull-jakauman ominaisuuksista	20
5.2	Kvantiilisekoitusten sovittuminen Weibull-aineistoihin	22
6	Sovellus matkapuhelinaineistoon	33
6.1	Aineisto	33
6.2	Tutkimuskysymys	34
6.3	Mallintaminen	34
6.3.1	Ikäryhmät	35
6.3.2	Alueet	38
6.4	Diagnostiikkaa	41
7	Yhteenvedo	45
	Viitteet	46
	Liite 1: Momentit	
	Liite 2: Aineisto	
	Liite 3: R-koodia	

1 Johdanto

Elinaikoja on perinteisesti mallinnettu parametrisilla malleilla, esimerkiksi Weibull-jakauman avulla. Parametriset elinaikamallit muodostavat joustavan kehikon mallintaa erilaisia ja erilaisilla tavoilla syntyneitä elinaika-aineistoja mm. riskifunktioiden avulla sekä tarjoavat keinon haluttaessa liittää malliin selittäviä muuttujia. (Jenkins, 2005, sivut 1 – 12)

Kvantiilisekoitusmallit ovat jatkuvia parametrisia malleja, jotka voidaan ymmärtää laajennuksina tai vaihtoehtoina yhden muuttujan parametrisille malleille. Parametrisissa malleissa on yleensä lähtökohtana mallintaminen parametrisoidun tiheysfunktion avulla. Kvantiilisekoitusmalleissa lähdetään jakauman kertymäfunktion käänteisfunktion eli kvantiilifunktion mallintamisesta. Tämä voidaan tehdä esimerkiksi mallintamalla kvantiilifunktio painotettuna summana erilaisista kvantiilifunktioista. Erityinen tyyppi kvantiilisekoituksia on polynomiset kvantiilisekoitukset, joiden komponentteina käytetään yhtä ns. pohjajakaumaa, joka on jokin jatkuva jakauma, ja tähän lisätään toisen asteen polynomi, jonka voi tulkita vinous- ja hui-pukkuuskomponentiksi. Kvantiilisekoitusmallit on esitetty artikkelissa (Karvanen, 2006).

Tämän tutkielman tarkoituksena on tutkia kvantiilisekoitusmallin käyttämistä parametrisena elinaikamallina. Miten kvantiilisekoitusmallit voidaan muotoilla elinaikojen mallinnukseen sopiviksi? Millaisia ominaisuuksia niillä on? Miten tällaiset mallit vertautuvat klassisiin parametrisiin elinaikamalleihin?

Tutkielmassa esitetään aluksi polynomisten kvantiilisekoitusten teoriaa ja parametrien estimointia. Esitetään tähän kolme soveltuvaa menetelmää. Suurimman uskottavuuden menetelmä (SU-menetelmä) soveltuu parametrien estimointiin, koska kyseessä on parametrinen malli. Estimaattorit eivät kuitenkaan ole yleensä saatavilla suljetussa muodossa, joten suuren otoskoon tapauksessa numeerisesti laskettavien estimaattien laskeminen voi olla laskennallisesti verrattaen vaativaa. Helposti juuri kvantiilisekoitusten parametrien estimointiin sovellettavissa oleva menetelmä on ns. L-momenttimenetelmä. Viimeisenä käsitellään kvantiilifunktioon sovellettu pienimmän neliösumman menetelmä (PNS-menetelmä).

Seuraavaksi on tutkittu kvantiilisekoitusten soveltamista elinaikatilanteeseen. Esitetään Karvasen johtaman normaali-polynomisen kvantiilisekoituksen käyttö elinaikamallina. Uusina tuloksina johdetaan sitten kaksi muuta elinaikojen mallintamiseen soveltuvaa kvantiilisekoitusta ja tarkastellaan näiden käyttäytymistä ja estimointia. Huomataan, että ensimmäinen näistä kahdesta on Weibull-jakauman laajennus.

Tutkielmassa tavoitteena on erityisesti tutkia, kuinka erilaiset kvantiilisekoitusmallit soveltuvat erityyppisiin elinaika-aineistoihin, ja kuinka ne toimivat klassisiin parametrisiin malleihin verrattuna. Tarkastellaan em. kolmen kvantiilisekoituksen soveltumista Weibull-jakautuneeseen aineistoon. Erityisesti soveltuvuus riskitiheysfunktion suhteen on kiinnostuksen kohteena.

Lopuksi tarkastellaan elinaika-aineistoa, jossa on havaittu suomalaisten käytössä olevien matkapuhelinten hankinnasta kuluneita aikoja. Tutkitaan, että eroavatko mediaanit eri-ikäisillä henkilöillä tai eri osissa maata. Aineisto on luonteeltaan väli-sensuroitua, joten mallintaminen parametrisen mallin avulla on luontevaa. Sovellaan tässä kolmea esiteltyä kvantiilisekoitusmallia.

2 Kvantiilisekoitukset

Esitellään tässä kvantiilisekoitusten idea yleisesti.

2.1 Yleinen kvantiilisekoitus

Kvantiilisekoitusmallilla tarkoitetaan yhden muuttujan parametrissa mallia, jossa jakauman kvantiilifunktio Q eli kertymäfunktion käänteisfunktio mallinnetaan lineaarikombinaationa useiden eri jakaumien kvantiilifunktioista Q_i määrittelemällä

$$Q(u) = \sum_{i=1}^k c_i Q_i(u), \quad 0 < u < 1, \quad (1)$$

jossa kertoimet c_i ovat reaalisia parametreja ja kvantiilifunktiot Q_i ovat tunnettuja.

Funktio Q on jonkin jakauman kvantiilifunktio vain silloin, kun $Q'(u) \geq 0$ kaikilla $u \in (0, 1)$. Tällöin Q määrää jakauman $F = Q^{-1}$, joka on kyseistä kvantiilisekoitusta vastaava kvantiilisekoitusjakauma. Huomattavaa on, että voi olla $c_i < 0$ joillekin $i = 1, \dots, k$, mutta ainakin yhden kertoimista on oltava ei-negatiivinen. Toisaalta ehto $c_i \geq 0$, kaikilla $i = 1, \dots, k$, on riittävä kvantiilifunktioehdon täyttymiselle. (Karvanen, 2006)

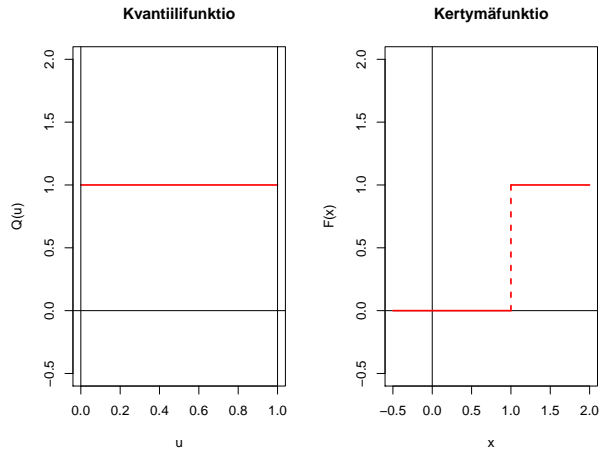
2.2 Polynominen kvantiilisekoitus

Tässä tarkastellaan eräitä kvantiilisekoitusten erikoistapauksia, polynomisia kvantiilisekoituksia. Ajatus jakauman kvantiilifunktion mallintamisesta polynomien avulla on esitetty ensi kertaa vuonna 1969 artikkelissa (Sillitto, 1969). Muun muassa artikkelissa (Malham ja Wiese, 2014) on approksimoitu summasatunnaismuuttujan kvantiilifunktiota polynomilla. Jakauman kvantiilifunktio mallinnetaan ns. pohjajakauman Q_0 ja toisen asteen polynomien summana, kuten artikkelissa (Karvanen, 2006), määrittelemällä

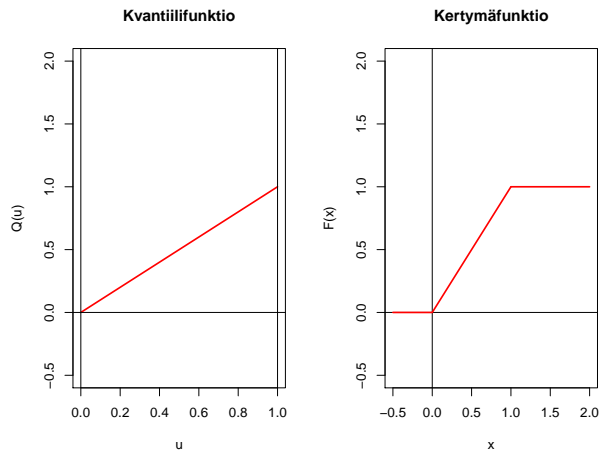
$$Q(u) = bQ_0(u) + a_2u^2 + a_1u + a_0. \quad (2)$$

Kyseessä on kvantiilisekoitus, koska toisen asteen polynomien termeistä kukin vastaa tietyn jakauman kvantiilifunktiota. Tämän muotoilun idea on lisätä jakaumaan Q_0 sijainti-, vinous- ja huipukkuuskomponentteja. Erityisesti kerroin a_0 vaikuttaa vain jakauman sijaintiin. Kertoimet a_1 ja a_2 vaikuttavat myös jakauman vinouteen ja huipukkuuteen. Polynomi voi olla korkeampaakin astetta, mutta uudet komponentit eivät yleensä paranna mallia kovin paljon, ja kertoimien tulkinta vaikeutuu niiden määrän kasvaessa. (Karvanen, 2006) Kvantiilisekoitusmallia on sovellettu esimerkiksi simulointikokeessa artikkelissa (Saarela et al., 2012).

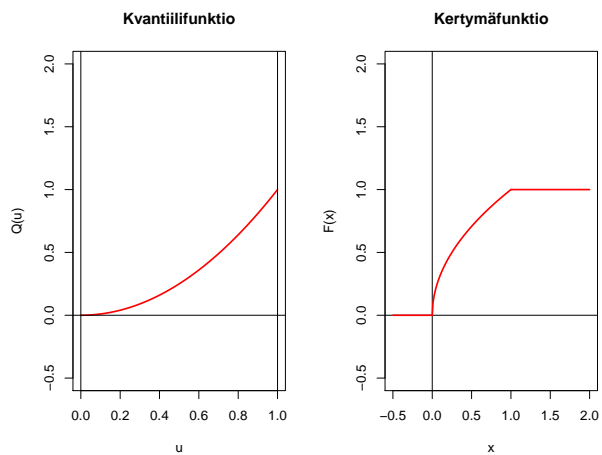
Kuvissa 1, 2 ja 3 on esitetty polynomien eri termejä vastaavien kvantiilifunktioiden käänteisfunktiot eli eri polynomikomponentteja vastaavat kertymäfunktiot.



Kuva 1: 1-komponenttia vastaavat kvantiili- ja kertymäfunktiot.



Kuva 2: u -komponenttia vastaavat kvantiili- ja kertymäfunktiot.



Kuva 3: u^2 -komponenttia vastaavat kvantiili- ja kertymäfunktiot.

Polynomikomponentit vaikuttavat siis kvantiilisekoitusjakaumaan rajatulla reaaliakselin välillä. Jos mallin halutaan käsittävän koko reaaliakselin, niin komponentin Q_0 on oltava sellainen. Tällaiselle jakaumalle $\lim_{u \rightarrow 1} Q_0(u) = \infty$ ja $\lim_{u \rightarrow 0} Q_0(u) = -\infty$. Tällöin $\lim_{u \rightarrow 1} Q'_0(u) = \infty$ ja $\lim_{u \rightarrow 0} Q'_0(u) = \infty$.

Perustelu. Jos olisi $\lim_{u \rightarrow 1} Q'_0(u) \leq L < \infty$ kaikilla $u \in (0, 1)$, niin esimerkiksi

$$Q_0(1) - Q_0\left(\frac{1}{2}\right) \leq L \left(1 - \frac{1}{2}\right)$$

differentiaalilaskennan väliarvolauseen nojalla välillä $\left(\frac{1}{2}, 1\right)$, joten olisi $Q_0(1) \leq \frac{L}{2} + Q_0\left(\frac{1}{2}\right) < \infty$, mikä ei voi päteä, koska $\lim_{u \rightarrow 1} Q_0(u) = \infty$. Arvolle $Q_0(0)$ saadaan samankaltainen päättely. \square

Koska polynomikomponenttien derivaatat ovat rajoitettuja, niin huomataan, että tässä tapauksessa on oltava $b \geq 0$, jotta kvantiilifunktioehto $0 \leq Q'(u) = bQ'_0(u) + 2a_2u + a_1$, $0 < u < 1$, pätsi. Jones (1992) käyttää kvantiilifunktion derivaatasta nimitystä kvantiilitiheysfunktio.

Tällaisilla kvantiilisekoitusjakaumilla on myös äärellisinä täsmälleen ne momentit kuin jakaumalla Q_0 . Erityisesti odotusarvo on äärellinen. Tämä voidaan muotoilla seuraavasti:

Olkoon $X \sim F$, joka on kvantiilisekoitusjakauma, jonka pohjajakauma on F_0 , ja $X_0 \sim F_0$. Tällöin, jos $\mathbb{E}(|X_0|^p) < \infty$, niin $\mathbb{E}(|X|^p) < \infty$, missä $p \geq 1$.

Perustelu. Olkoon $X \sim F$ ja $u_0 = F(0)$. Merkitään $|a_2| + |a_1| + |a_0| = c$. Tällöin, jos $u \geq u_0$, niin

$$\begin{aligned} 0 \leq Q(u) &\leq |Q(u)| = |bQ_0(u) + a_2u^2 + a_1u + a_0| \\ &\leq |bQ_0(u)| + |a_2u^2 + a_1u + a_0| \leq bQ_0(u) + |a_2| + |a_1| + |a_0| \\ &= bQ_0(u) + c = Q_A(u). \end{aligned}$$

Tämä Q_A on muunnoksen $A = bX_0 + c$ kvantiilifunktio. Tällä muuttujalla on äärellisinä samat momentit kuin jakaumalla F_0 . (Liite 1) Nyt kvantiilifunktioita Q ja Q_A vastaaville kertymäfunktioille pätee $F_A(x) \leq F(x)$, kun $x \geq 0$. Vastaavasti arvoille $u_0 \leq 0$ saadaan

$$\begin{aligned} 0 \geq Q(u) &\geq -|Q(u)| = -|bQ_0(u) + a_2u^2 + a_1u + a_0| \\ &\geq -|bQ_0(u)| - |a_2u^2 + a_1u + a_0| \geq bQ_0(u) - |a_2| - |a_1| - |a_0| \\ &= bQ_0(u) - c = Q_S(u). \end{aligned}$$

Tämä Q_S on muunnoksen $S = bX_0 - c$ kvantiilifunktio. Tällä muuttujalla on äärellisinä samat momentit kuin jakaumalla F_0 . Nyt kvantiilifunktioita Q ja Q_S vastaaville kertymäfunktioille pätee $F_S(x) \geq F(x)$, kun $x < 0$. Kuvassa 4 on hahmoteltu näitä kertymäfunktioita.

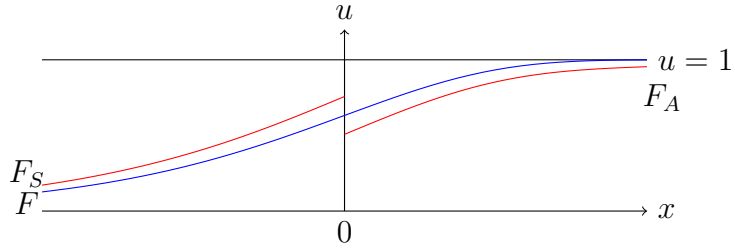
Olkoot X^+ ja X^- muuttujan X positiivi- ja negatiiviosat. Tällöin muuttujan $|X|$ kertymäfunktio

$$F_{|X|}(x) = P(|X| < x) = P(-x < X < x) = F(x) - F(-x) \geq F_A(x) - F_S(-x).$$

Tämän avulla nähdään, että

$$\begin{aligned}\mathbb{E}(|X|^p) &= p \int_0^\infty x^{p-1}(1 - F_{|X|}(x))dx \leq p \int_0^\infty x^{p-1}(1 - F_A(x) + F_S(-x))dx \\ &= p \int_0^\infty x^{p-1}(1 - F_A(x))dx + p \int_0^\infty x^{p-1}F_S(-x)dx \\ &= \mathbb{E}([A^+]^p) + \mathbb{E}([S^-]^p) < \infty,\end{aligned}$$

jos $\mathbb{E}(|X_0|^p) < \infty$.



Kuva 4: Hahmotelmaa päättelyssä käytetyistä kertymäfunktioista.

□

3 Kvantiilisekoitusten parametrien estimointi

Parametrisen mallin käyttöön liittyy oleellisena osana mallin parametrien estimointi otoksesta y_i , $i = 1, \dots, n$, jonka oletetaan noudattavan jotain tämän jakauma-perheen jakaumaa. Esitellään tässä luvussa kolme menetelmää kvantiilisekoitusten parametrien estimointiin.

3.1 Suurimman uskottavuuden menetelmä

Parametrisena mallina kvantiilisekoitusmallin parametrit voidaan estimoida tavalliseen tapaan SU-menetelmällä. Jos oletetaan, että $Q(u) = bQ_0(u) + a_2u^2 + a_1u + a_0$, niin havaintoa y_i , $i = 1, \dots, n$, vastaava kertymäfunktion arvo u_i löydetään yhtälön

$$y_i = Q(u_i) = bQ_0(u_i) + a_2u_i^2 + a_1u_i + a_0 \quad (3)$$

ratkaisuna. Huomataan, että oikea puoli on arvon u kasvava funktio, joten ratkaisu on helppo löytää esim. puolitushaulla. Mallin mukainen tiheysfunktio voidaan kirjoittaa arvon u_i avulla

$$f(y_i|b, a_2, a_1, a_0) = f(Q(u_i)|b, a_2, a_1, a_0) = \frac{1}{bQ_0'(u_i) + 2a_2u_i + a_1}, \quad (4)$$

koska $Q'(u) = bQ_0'(u) + 2a_2u + a_1 \geq 0$ kaikilla $u \in (0, 1)$. (Karvanen (2006), Jones (1992))

Perustelu. Olkoon $Y \sim F$. Jos $U \sim Tas(0, 1)$, niin muuttuja Y on samoin jakautunut kuin muunnos $Q(U) = F^{-1}(U)$, joten muuttujan Y tiheysfunktio on

$$\begin{aligned} f_Y(y) &= f_{Q(U)}(Q(u)) = f_U(Q^{-1}(y)) \left| \frac{d}{dy} Q^{-1}(y) \right| \\ &= \frac{f_U(u)}{\left| \frac{d}{du} Q(u) \right|} = \frac{f_U(u)}{Q'(u)} = \frac{1}{bQ_0'(u) + 2a_2u + a_1}, \quad \text{missä } y = Q(u). \end{aligned}$$

□

Tällöin otoksen y_1, \dots, y_n (i.i.d.) uskottavuusfunktio voidaan kirjoittaa:

$$L(b, a_2, a_1, a_0|y_1, \dots, y_n) = \prod_{i=1}^n f(y_i|b, a_2, a_1, a_0). \quad (5)$$

Vastaavasti, jos osa havainnoista on sensuroitunut, eli niistä tunnetaan vain väli $y_i \in (c_i^l, c_i^u)$, $m + 1 \leq i \leq n$, voidaan uskottavuusfunktio kirjoittaa:

$$\begin{aligned} L(b, a_2, a_1, a_0|y_1, \dots, y_n) &= \\ &= \prod_{i=1}^m f(y_i|b, a_2, a_1, a_0) \cdot \prod_{i=m+1}^n \left[F(c_i^u|b, a_2, a_1, a_0) - F(c_i^l|b, a_2, a_1, a_0) \right]. \end{aligned} \quad (6)$$

Maksimoimalla tämä uskottavuusfunktio löydetään suurimman uskottavuuden ratkaisu $(\hat{b}, \hat{a}_2, \hat{a}_1, \hat{a}_0)^T$ ja siten myös kvantiilisekoitusta vastaava estimoitu jakauma. Tämä joudutaan käytännössä aina tekemään numeerisesti ja voi olla raskaahko laskea otoskoon n ollessa suuri.

3.2 L-momenttimenetelmä

Vaihtoehtoinen tapa estimoida parametrit on ns. L-momenttimenetelmä. (Hosking, 1990) Menetelmä muistuttaa tavallista momenttimenetelmää, jossa otoksen tietyt momentit asetetaan samoiksi vastaavien mallin teoreettisten momenttien kanssa. Vastaavasti L-momenttimenetelmässä asetetaan samoiksi otoksesta laskettuja L-momentteja mallijakauman L-momenttien kanssa.

3.2.1 L-momentit

L-momentit (lineaariset momentit) ovat perinteisten momenttien kaltainen joukko jakauman tunnuslukuja, jotka kuvaavat jakauman sijaintia, vaihtelua, vinoutta, huipukkuutta ym. Ne ovat järjestettyihin otoksiin perustuvia tunnuslukuja. Hosking (1990) esittelee L-momenttien teoriaa ja johtaa teoreettiset L-momentit seuraavalla tavalla:

Olkoot $Z_i \sim F$, $i = 1, \dots, n$, jakauman F mukaan jakautuneita riippumattomia satunnaismuuttujia. Olkoot tällöin $Z_{1:n}, \dots, Z_{n:n}$ vastaavat nousevaan järjestykseen järjestetyt muuttujien arvot. Tällöin $Z_{k:n}$, $k = 1, \dots, n$, on siis muuttujajoukon k . pienintä arvoa kuvaava satunnaismuuttuja. Tällöin asteen r L-momentti määritellään

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EZ_{(r-k):r}, \quad (7)$$

missä

$$\begin{aligned} EZ_{(r-j):r} &= \frac{r!}{(j-1)!(r-j)!} \int [F(z)]^{j-1} [1-F(z)]^{r-j} z dF(z) \\ &= \frac{r!}{(j-1)!(r-j)!} \int_0^1 u^{j-1} (1-u)^{r-j} Q_F(u) du. \end{aligned}$$

Tällöin erityisesti neljä ensimmäistä L-momenttia voidaan jakauman F kvanttilifunktion Q_F avulla kirjoittaa:

$$\begin{aligned} \lambda_1 &= \frac{1}{1} E(Z_{1:1}) = \int_0^1 Q_F(u) du, \\ \lambda_2 &= \frac{1}{2} E(Z_{2:2} - Z_{1:2}) = \int_0^1 (2u-1) Q_F(u) du, \\ \lambda_3 &= \frac{1}{3} E(Z_{3:3} - 2Z_{2:3} + Z_{1:3}) = \int_0^1 (6u^2 - 6u + 1) Q_F(u) du \text{ ja} \\ \lambda_4 &= \frac{1}{4} E(Z_{4:4} - 3Z_{3:4} + 3Z_{2:4} - Z_{1:4}) = \int_0^1 (20u^3 - 30u^2 + 12u - 1) Q_F(u) du. \end{aligned} \quad (8)$$

Vastaavasti kuin tavalliset momentit, nämä neljä ensimmäistä L-momenttia kuvaavat jakauman sijaintia, vaihtelua, vinoutta ja huipukkuutta. L-momenttien olemassaolon toteamiseen riittää, että jakauman odotusarvo on äärellinen.

L-momenttien estimointi otoksesta

L-momenttien suora estimointi otoksesta siten, että käydään läpi kaikki r havainnon aliotokset ja keskiarvoistetaan ne, on laskennallisesti vaativaa otoksen ollessa suuri. Hosking (1990) esittää tehokkaamman kaavan estimoida L-momentit otoksesta suoraan lineaarikombinaatioina, joiden kertoimet ovat tietynlaisia alternoivia summia. Elamir ja Seheult (2003) esittävät kaavan seuraavasti:

Olkoon $y_{1:n}, \dots, y_{n:n}$ järjestetty otos kasvavassa järjestyksessä. Tällöin saadaan L-momentin λ_r estimaatti \bar{l}_r seuraavalla kaavalla:

$$\bar{l}_r = r^{-1} \sum_{i=1}^n \left(\frac{\sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \binom{i-1}{r-1-k} \binom{n-i}{k}}{\binom{n}{r}} \right) y_{i:n}. \quad (9)$$

Jos kaavan binomikertoimissa $\binom{n}{k}$ ei päde $0 \leq k \leq n$, niin $\binom{n}{k} = 0$. Tämä on implementoitu R-kirjastossa *Lmoments*. (Karvanen, 2011).

L-momenttien estimointi on varsin robustia myös pienillä aineistoilla. Elamir ja Seheult (2004) esittävät estimaattoreille (9) eksaktin varianssi – kovarianssimatriisin estimaattorin $\tilde{\Sigma}_l$. Tämä on implementoitu R-kirjastossa *nsRFA*. (Viglion, 2014)

3.2.2 Parametrien estimointi L-momenttien avulla

Avainasemassa on huomio, että kvantiilisekoitusmallin tapauksessa voidaan kirjoittaa:

$$\lambda_r(Q) = \lambda_r\left(\sum_{i=1}^k c_i Q_i\right) = \sum_{i=1}^k c_i \lambda_r(Q_i), \quad (10)$$

joka seuraa suoraan esimerkiksi yhtälöistä (8). Tapauksessa $r = 2$ nähdään, että

$$\begin{aligned} \lambda_2(Q) &= \lambda_2\left(\sum_{i=1}^k c_i Q_i\right) = \int_0^1 (2u - 1) \left[\sum_{i=1}^k c_i Q_i(u) \right] du \\ &= \sum_{i=1}^k c_i \int_0^1 (2u - 1) Q_i(u) du = \sum_{i=1}^k c_i \lambda_2(Q_i). \end{aligned}$$

Muut L-momentit käyttäytyvät vastaavasti. □

Nämä yksinkertaisten jakaumien L-momentit $\lambda_r(Q_i)$ ovat yleensä helposti laskettavissa, ja erityisesti polynomin komponenteille neljä ensimmäistä L-momenttia ovat seuraavat:

$$\begin{array}{llllll} \lambda_1(Q_1) & = 1 & \lambda_2(Q_1) & = 0 & \lambda_3(Q_1) & = 0 & \lambda_4(Q_1) & = 0 \\ \lambda_1(Q_u) & = \frac{1}{2} & \lambda_2(Q_u) & = \frac{1}{6} & \lambda_3(Q_u) & = 0 & \lambda_4(Q_u) & = 0 \\ \lambda_1(Q_{u^2}) & = \frac{1}{3} & \lambda_2(Q_{u^2}) & = \frac{1}{6} & \lambda_3(Q_{u^2}) & = \frac{1}{30} & \lambda_4(Q_{u^2}) & = 0. \end{array} \quad (11)$$

Jakauman parametrien estimointi tapahtuu nyt asettamalla samoiksi otoksesta estimoituja L-momentteja jakauman teoreettisten L-momenttien kanssa. (Gilchrist, 2000, sivut 74 – 79, 194) Koska mallissa on neljä parametria, niin ratkaisun saamiseksi on kiinnitettävä neljä momenttia samoiksi. Luonnollinen valinta on kiinnittää neljä ensimmäistä L-momenttia siten, että

$$\bar{l}_r = \lambda_r(bQ_0 + a_2Q_{u^2} + a_1Q_u + a_0Q_1), \quad r = 1, \dots, 4. \quad (12)$$

Nyt L-momenttiestimaatti $(\tilde{b}, \tilde{a}_2, \tilde{a}_1, \tilde{a}_0)^T$ saadaan yhtälöryhmän

$$\begin{pmatrix} \bar{l}_1 \\ \bar{l}_2 \\ \bar{l}_3 \\ \bar{l}_4 \end{pmatrix} = \begin{pmatrix} \lambda_1(Q_0) & \frac{1}{3} & \frac{1}{2} & 1 \\ \lambda_2(Q_0) & \frac{1}{6} & \frac{1}{6} & 0 \\ \lambda_3(Q_0) & \frac{1}{30} & 0 & 0 \\ \lambda_4(Q_0) & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b \\ a_2 \\ a_1 \\ a_0 \end{pmatrix} = A \begin{pmatrix} b \\ a_2 \\ a_1 \\ a_0 \end{pmatrix} \quad (13)$$

ratkaisuna

$$\begin{aligned} \tilde{b} &= \frac{1}{\lambda_4(Q_0)} \bar{l}_4, \\ \tilde{a}_2 &= 30\bar{l}_3 - 30 \frac{\lambda_3(Q_0)}{\lambda_4(Q_0)} \bar{l}_4, \\ \tilde{a}_1 &= 6\bar{l}_2 - 30\bar{l}_3 + \frac{30\lambda_3(Q_0) - 6\lambda_2(Q_0)}{\lambda_4(Q_0)} \bar{l}_4 \text{ ja} \\ \tilde{a}_0 &= \bar{l}_1 - 3\bar{l}_2 + 5\bar{l}_3 + \frac{-5\lambda_3(Q_0) + 3\lambda_2(Q_0) - \lambda_1(Q_0)}{\lambda_4(Q_0)} \bar{l}_4. \end{aligned} \quad (14)$$

L-momenttiratkaisu on hyödyllinen sellaisenaan, mutta erityisesti sitä voi käyttää SU-menetelmän lähtöarvona uskottavuusfunktion numeerisessa maksimoinnissa. Toisaalta L-momenttien robustisuus pienillä otoksilla periytyy myös itse estimaattoreille $(\tilde{b}, \tilde{a}_2, \tilde{a}_1, \tilde{a}_0)^T$, mikä voi olla tavoiteltava hyöty. Parametriestimaattoreille saadaan varianssi – kovarianssimatriisi L-momenttien varianssi – kovarianssimatriisin avulla huomaamalla, että $\tilde{\Sigma}_{(\tilde{b}, \tilde{a}_2, \tilde{a}_1, \tilde{a}_0)^T} = A^{-1} \tilde{\Sigma}_l (A^{-1})^T$. (Karvanen, 2006)

3.3 Kvantiilifunktion pienimmän neliösumman menetelmä

Kvantiilisekoituksen parametrit on myös mahdollista estimoida minimoimalla otoksen empiirisen kvantiilifunktion ja mallikvantiilifunktion erotuksen neliösumma pisteissä $\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}$ ja 1, jossa n on otoskoko. (Gilchrist, 2000, sivut 193 – 222)

Estimaatti

$$\begin{aligned} (\check{b}, \check{a}_2, \check{a}_1, \check{a}_0)^T &= \arg \min_{b, a_2, a_1, a_0} \sum_{i=1}^n (Q_n(\frac{i}{n}) - Q(\frac{i}{n}))^2 \\ &= \arg \min_{b, a_2, a_1, a_0} \sum_{i=1}^n (y_{i:n} - Q(\frac{i}{n}))^2, \end{aligned} \quad (15)$$

jossa Q_n on otoksen empiirinen kvantiilifunktio ja Q on kvantiilisekoitusmalli.

4 Polynominen kvantiilisekoitusmalli elinajoille

Tässä luvussa tarkastellaan, kuinka kvantiilisekoitusmalleja voidaan soveltaa elinajojen mallintamiseen. Muodostetaan kolme erilaista kvantiilisekoitusmallia tähän tarkoitukseen ja tutkitaan niiden käyttäytymistä.

4.1 Lyhyesti elinajoista

Elinajaksi kutsutaan sellaista muuttujaa, joka voi saada vain ei-negatiivisia arvoja. Kaikki muuttujat, joilla on jokin kiinteä äärellinen alaraja, voidaan itse asiassa tulkita tällaisiksi. Nimitys elinaika tulee siitä, että tällainen muuttuja voidaan yleensä tulkita jonkinlaiseksi ajaksi, joka kuuluu alkuhetki nolasta johonkin yksilölle tapahtuvaan tapahtumaan, kuten eliön kuolemaan tai laitteen rikkoutumiseen.

Tällaisesta tulkinnasta tapahtumaan kuluvana aikana saadaan elinaikajakaumille käyttöön kätevä funktio, joka kuvaa tapahtumatiheyttä. Tämä elinajojen jakaumaa kuvaava funktio on sen ns. riskitiheysfunktio. Muuttujan X riskitiheysfunktio

$$R_X(x) = \frac{f(x)}{1 - F(x)}, \quad x > 0. \quad (16)$$

Tässä F on jakauman kertymä- ja f tiheysfunktio. Riskitiheysfunktio kuvaa muuttujan X todennäköisyystiheyttä saada arvo kyseisellä hetkellä x ehdolla, että $X \geq x$, eli se on yksilön kuolemistodennäköisyystiheys hetkellä x ehdolla, että yksilö on selviytynyt hetkeen x asti. Esimerkiksi laskeva riski vastaa paksuhäntäistä jakaumaa, jossa muuttujan X odotettavissa oleva elinaika tulee suuremmaksi muuttujan x kasvaessa.

4.2 Kvantiilisekoitusmallin soveltaminen elinaikoihin

Tapaukseen, jossa aineisto on tyyppiä $x_i > 0$, $i = 1, \dots, n$, ei kvantiilisekoitusmalli sovi sellaisenaan. Kuitenkin, jos alkuperäiselle aineistolle tehdään muunnos

$$y_i = \log(x_i), \quad i = 1, \dots, n, \quad (17)$$

niin muunnoksen kantajaksi tulee koko reaaliakseli, jolloin tähän muunnettuun aineistoon voidaan sovittaa kvantiilisekoitusmalli. Kaikissa seuraavissa tapauksissa aineisto muunnetaan kaavalla (17). Haluttaessa tarkastella mallin sovitetta alkuperäisessä skaalassa täytyy tiheysfunktio muuntaa takaisin käänteismuunnoksella. Logaritmimuunnos on aidosti kasvava, joten tiheysfunktio alkuperäisessä X -skaalassa saadaan muunnoksella

$$f_X(x) = f_Y(\log(x)) \frac{1}{x}, \quad x > 0. \quad (18)$$

Tällä tavalla muunnetulla kvantiilisekoituksella on samat momentit äärellisinä kuin muuttujalla e^{bX_0} , jossa $X_0 \sim F_0$. Eksponentoidun kvantiilisekoituksen momenttien olemassaolo riippuu siis myös parametrasta b , mikä nähdään seuraavasti:

Olkoon $X \sim F$, joka on kvantiilisekoitusjakauma, jonka pohjajakauma on F_0 , ja $X_0 \sim F_0$. Tällöin, jos $\mathbb{E}(|e^{bX_0}|^p) < \infty$, niin $\mathbb{E}(|e^X|^p) < \infty$, missä $p \geq 1$.

Perustelu. Käytetään apuna kappaleen 2.2 merkintöjä ja tuloksia. Todetaan, että

$$F_{e^X}(x) = P(e^X \leq x) = P(X \leq \log(x)) = F(\log(x)),$$

ja vastaava pätee muuttujalle A , joten kertymäfunktiolle F_{e^X} saadaan arvio

$$F_{e^X}(x) = F(\log(x)) \geq F_A(\log(x)) = F_{e^A}(x),$$

kun $x \geq 1$. Muuttujalla $e^A = e^{bX_0+c}$ on äärellisinä samat momentit kuin muuttujalla e^{bX_0} . (Liite 1) Siten nähdään, että

$$\begin{aligned} \mathbb{E}(|e^X|^p) &= p \int_0^\infty x^{p-1}(1 - F_{e^X}(x))dx \leq p \int_0^1 x^{p-1}(1 - F_{e^X}(x))dx \\ &\quad + p \int_1^\infty x^{p-1}(1 - F_{e^A}(x))dx \leq p \int_0^1 x^{p-1}(1 - F_{e^X}(x))dx + \mathbb{E}(|e^A|^p) < \infty, \end{aligned}$$

jos $\mathbb{E}(|e^{bX_0}|^p) < \infty$. □

4.3 Lognormaali-polynominen kvantiilisekoitus

Artikkelissa (Karvanen, 2006) on tutkittu normaali-polynomista kvantiilisekoitusta, jossa Q_0 on $N(0, 1)$ -jakauman kvantiilifunktio. Sovelletaan tätä kvantiilisekoitusmallia logaritmoituun aineistoon. Alkuperäisen aineiston näkökulmasta tätä voidaan nimittää lognormaali-polynomiseksi kvantiilisekoitukseksi. On huomattavaa, että nimi ei ole täysin yhtenevä normaali-polynomisen sekoituksen kanssa, sillä se ei ole sama asia kuin kvantiilisekoitus, jossa lognormaaliin jakaumaan yhdistettäisiin polynomi-komponentteja. Sen sijaan polynomikomponentit ovat tässä mukana tietyllä tavalla eksponentoidussa muodossaan.

Olkoon otos x_1, \dots, x_n , josta saadaan kaavalla (17) logaritmoitu otos y_1, \dots, y_n . Estimointi voidaan tehdä kolmella luvussa 3 esitetyllä tavalla.

L-momenttiestimaattia varten otoksesta voidaan laskea kaavalla (9) L-momenttien estimaatit $\bar{l}_1, \bar{l}_2, \bar{l}_3, \bar{l}_4$. Itse estimointiyhtälöryhmä (13) saa lognormaali-polynomisen kvantiilisekoituksen tapauksessa muodon:

$$\begin{pmatrix} \bar{l}_1 \\ \bar{l}_2 \\ \bar{l}_3 \\ \bar{l}_4 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{\sqrt{\pi}} & \frac{1}{30} & \frac{1}{6} & 1 \\ \frac{1}{\sqrt{\pi}} & \frac{6}{30} & \frac{1}{6} & 0 & 0 \\ 0 & \frac{30}{\pi} \arctan(\sqrt{2}) - 9 & 0 & 0 & 0 \\ \frac{30}{\pi} \arctan(\sqrt{2}) - 9 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b \\ a_2 \\ a_1 \\ a_0 \end{pmatrix}. \quad (19)$$

Tässä matriisin ensimmäisen sarakkeen arvot ovat $N(0, 1)$ -jakauman L-momentteja, jotka on laskettu kaavojen (8) avulla.

Estimoidun kvantiilisekoitusjakauman tiheysfunktio voidaan kirjoittaa muodossa:

$$\begin{aligned} f_Y(y) &= f_Y(Q(u)|\tilde{b}, \tilde{a}_2, \tilde{a}_1, \tilde{a}_0) = \frac{1}{\tilde{b}Q_0'(u) + 2\tilde{a}_2u + \tilde{a}_1} \\ &= \frac{1}{\frac{\tilde{b}}{F_0'(Q_0(u))} + 2\tilde{a}_2u + \tilde{a}_1} = \frac{1}{\frac{\tilde{b}}{f_0(Q_0(u))} + 2\tilde{a}_2u + \tilde{a}_1}, \end{aligned} \quad (20)$$

missä $y = Q(u)$. Kaavasta (18) saadaan tästä alkuperäisen skaalan tiheysfunktio. Tämän kvantiilisekoituksen funktiot ja parametrien estimointi on implementoitu R-kirjastossa *Lmoments*. (Karvanen, 2011)
 SU- ja PNS-estimaatit saadaan edellä esitetyllä tavalla.

4.3.1 Simulointiesimerkki

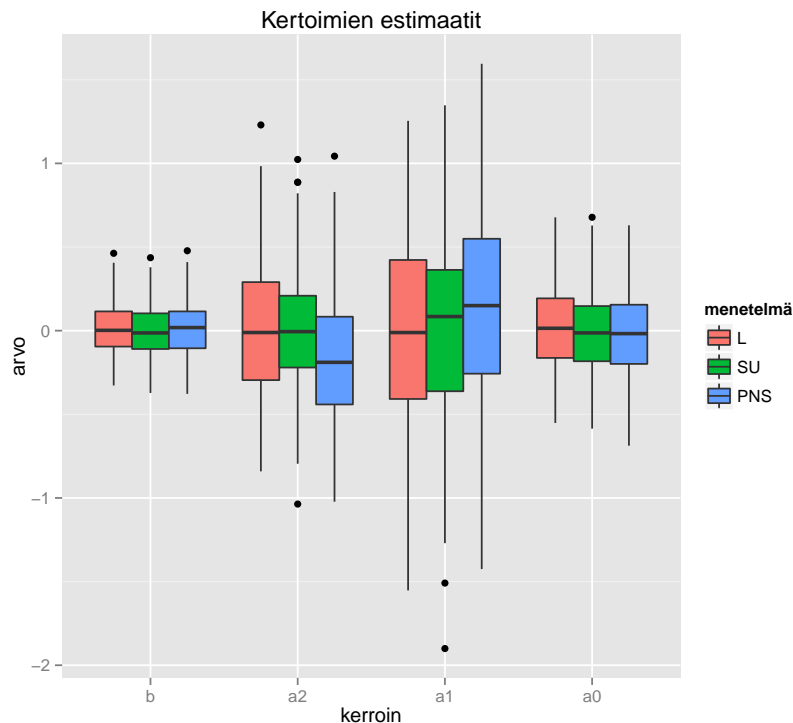
Tarkastellaan esiteltyjen kolmen estimointimenetelmän käyttäytymistä simuloimalla lognormaali-polynomisesta kvantiilisekoitusjakaumasta 200 kertaa 200 havainnon suuruinen otos. Käytetään näin pientä määrää, koska myöhemmässä simuloinnissa joudutaan tekemään sama laskennallisesti raskaammalle mallille. Käytetään jakaumaa, jonka parametrien arvot ovat $b = 1$, $a_2 = -0.5$, $a_1 = -1$ ja $a_0 = 0$. Kyseessä on siis jakauma, jonka kvantiilifunktio on

$$Q(u) = Q_{N(0,1)}(u) - 0.5u^2 - u, \quad 0 < u < 1. \quad (21)$$

Estimoidaan kvantiilisekoituksen parametrit kolmella menetelmällä kullekin simulointiotokselle ja tarkastellaan näiden parametriestimaattien jakaumaa. Verrataan lisäksi sovitteita simulointijakaumaan käyttämällä mittana Kolmogorov – Smirnov -etäisyyttä. Kolmogorov – Smirnov -etäisyys on kahden yksiulotteisen jakauman eroa kuvaava suure. Jakaumien F_1 ja F_2 välille se voidaan määritellä seuraavasti:

$$D_{KS}(F_1, F_2) = \sup_{x \in S} |F_1(x) - F_2(x)|, \quad (22)$$

jossa S tarkoittaa jakauman kantajaa.



Kuva 5: Lognormaali-polynomisesta kvantiilisekoitusjakaumasta poimituille otoksille eri menetelmällä laskettujen oikean arvon suhteen keskitettyjen parametriestimaattien viiksilaatikkokuvaajat.

Näyttää, että SU-menetelmä toimii näistä parhaiten, koska se on keskimäärin lähimpänä oikeaa, ja estimaattien vaihtelu on pienintä. L-momenttimenetelmällä saadaan lähes yhtä hyviä tuloksia, mutta PNS-menetelmän estimaateissa on jonkin verran harhaa, vaikka sekin menetelmä toimii varsin hyvin tässä mallissa. Erityisesti kertoimet a_2 ja a_1 estimoituvat hieman heikosti. Näiden kertoimien vaihtelu on itse asiassa kaikilla kolmella menetelmällä varsin isoa, koska molemmat kuvaavat hyvin samanlaista vaikutusta, mikä tuo yksittäisen kertoimen estimointiin epävarmuutta.

Taulukko 1: Edellisistä 200 otoksesta eri menetelmillä saatujen parametrivektoreiden keskiarvot ja niitä vastaavien sovitteiden ja simulointijakauman välisten Kolmogorov – Smirnov -tunnuslukujen keskiarvot sekä kaikkien näiden keskihajonnat.

Parametri	L-momenttimenetelmä	SU-menetelmä	PNS-menetelmä
$b = 1$	1.010 (0.144)	0.993 (0.142)	1.013 (0.154)
$a_2 = -0.5$	-0.501 (0.401)	-0.492 (0.340)	-0.669 (0.384)
$a_1 = -1$	-1.031 (0.575)	-0.989 (0.543)	-0.869 (0.573)
$a_0 = 0$	0.016 (0.233)	-0.008 (0.228)	-0.020 (0.246)
K-S	0.040 (0.017)	0.037 (0.017)	0.041 (0.018)

Taulukossa 1 on estimoidun parametrivektorin keskiarvot ja keskihajonnat menetelmittäin. Huomataan, että L-momentti- ja PNS-estimaatit ovat hyvin lähellä toisiaan ja poikkeavat jonkin verran SU-estimaatista, joka vaikuttaa olevan lähimpänä oikeaa.

4.4 Eksponentti-polynominen kvantiilisekoitus

Luonnollinen valinta pohjajakaumaksi Q_0 elinaikatilanteessa voisi olla eksponenttijakauma. Tämä yhdistettynä vinouskomponentteihin voisi olla toimiva tapa mallintaa elinaikoja kvantiilisekoitusten avulla. Siirryttäessä logaritmoituun aineistoon on kuitenkin luonnollista logaritmoida myös pohjajakaumaksi tarkoitettu eksponenttijakauma. Siis nyt $Q_0(u) = Q_{\log(X)}(u) = \log\left(\log\left(\frac{1}{1-u}\right)\right)$, missä $X \sim \text{Exp}(1)$.

Perustelu.

$$\begin{aligned}
 F_X(x) &= 1 - e^{-x}, \text{ jolloin} \\
 F_{\log(X)}(x) &= P(\log(X) \leq x) = P(X \leq e^x) = F_X(e^x) \\
 &= 1 - e^{-e^x}, \text{ jonka käänteisfunktio} \\
 F_{\log(X)}^{-1}(u) &= Q_{\log(X)}(u) = \log\left(\log\left(\frac{1}{1-u}\right)\right).
 \end{aligned}$$

□

Käytetään tästä nimitystä eksponentti-polynominen kvantiilisekoitus.

Olkoon logaritmoitu otos y_1, \dots, y_n . Estimointi voidaan tehdä kolmella luvussa 3 esitetyllä tavalla.

L-momenttiestimaattia varten otoksesta voidaan laskea kaavalla (9) L-momenttien estimaatit $\bar{l}_1, \bar{l}_2, \bar{l}_3, \bar{l}_4$. Itse estimointiyhtälöryhmä (13) saa eksponentti-polynomisen kvantiilisekoituksen tapauksessa muodon:

$$\begin{pmatrix} \bar{l}_1 \\ \bar{l}_2 \\ \bar{l}_3 \\ \bar{l}_4 \end{pmatrix} = \begin{pmatrix} \lambda_{1,\log X} & \frac{1}{3} & \frac{1}{2} & 1 \\ \lambda_{2,\log X} & \frac{1}{6} & \frac{1}{6} & 0 \\ \lambda_{3,\log X} & \frac{1}{30} & 0 & 0 \\ \lambda_{4,\log X} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b \\ a_2 \\ a_1 \\ a_0 \end{pmatrix}, \quad (23)$$

jossa $\lambda_{i,\log(X)}$, $i = 1, \dots, 4$, saadaan laskettua numeerisesti kaavasta (8) käyttäen kvantiilifunktiota $Q_{\log(X)}$:

$$\lambda_{1,\log X} = -0.577, \quad \lambda_{2,\log X} = 0.693, \quad \lambda_{3,\log X} = -0.118 \quad \text{ja} \quad \lambda_{4,\log X} = 0.104.$$

Estimoidun jakauman tiheysfunktio saadaan vastaavasti, kuten kaavassa (20), ja se saa logeksponenttisen pohjafunktion tilanteessa muodon:

$$\begin{aligned} f_Y(y) &= f_Y(Q(u)|\tilde{b}, \tilde{a}_2, \tilde{a}_1, \tilde{a}_0) \\ &= \frac{1}{\tilde{b} \frac{d}{du} \left[\log \left(\log \left(\frac{1}{1-u} \right) \right) \right] + 2\tilde{a}_2 u + \tilde{a}_1} \\ &= \frac{1}{\frac{-\tilde{b}}{(1-u)\log(1-u)} + 2\tilde{a}_2 u + \tilde{a}_1}. \end{aligned} \quad (24)$$

Kaavalla (18) saadaan tästä alkuperäisen skaalan tiheysfunktio. Jakauman funktiot ja parametrien estimointi L-momenttien avulla ovat R-koodissa liitteessä 3.

SU- ja PNS-estimaatit saadaan edellä esitetyllä tavalla.

4.4.1 Yhteys Weibull-jakaumaan

Olkoon $X \sim \text{Weibull}(\beta, k)$, jolloin muuttujan X kertymäfunktio on

$$F_{X_{\beta,k}}(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^k}, & \text{kun } x \geq 0 \\ 0, & \text{kun } x < 0. \end{cases} \quad (25)$$

Weibull-jakautuneen muuttujan logaritmin $\log(X)$ kvantiilifunktio voidaan tällöin kirjoittaa:

$$Q_{\log X_{\beta,k}}(u) = \frac{\log \left(\log \left(\frac{1}{1-u} \right) \right)}{k} + \log(\beta), \quad 0 < u < 1. \quad (26)$$

Perustelu.

$$\begin{aligned} F_{\log X_{\beta,k}}(x) &= P(\log(X) \leq x) = P(X \leq e^x) = F_{X_{\beta,k}}(e^x) \\ &= 1 - e^{-\left(\frac{e^x}{\beta}\right)^k}, \quad \text{jonka käänteisfunktio } Q_0 \text{ saadaan ratkaisemalla} \end{aligned}$$

$$u = 1 - e^{-\left(\frac{e^{Q_0(u)}}{\beta}\right)^k}$$

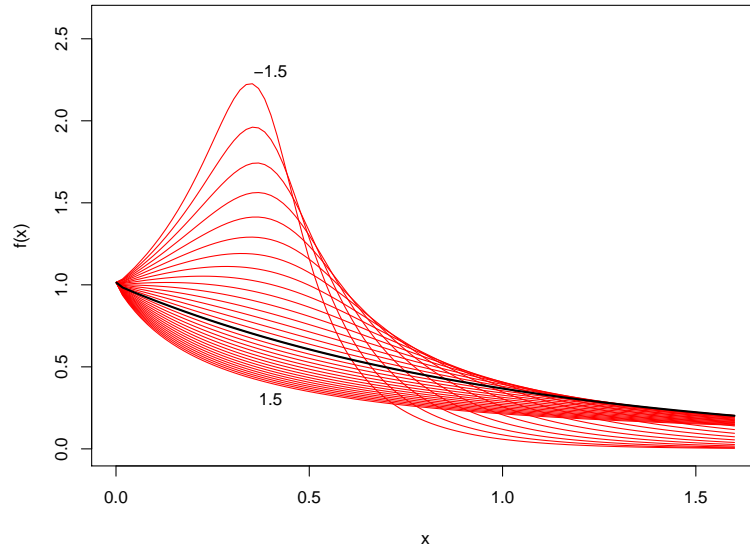
$$\left(\frac{e^{Q_0(u)}}{\beta}\right)^k = -\log(1-u)$$

$$k(Q_0(u) - \log(\beta)) = \log(-\log(1-u))$$

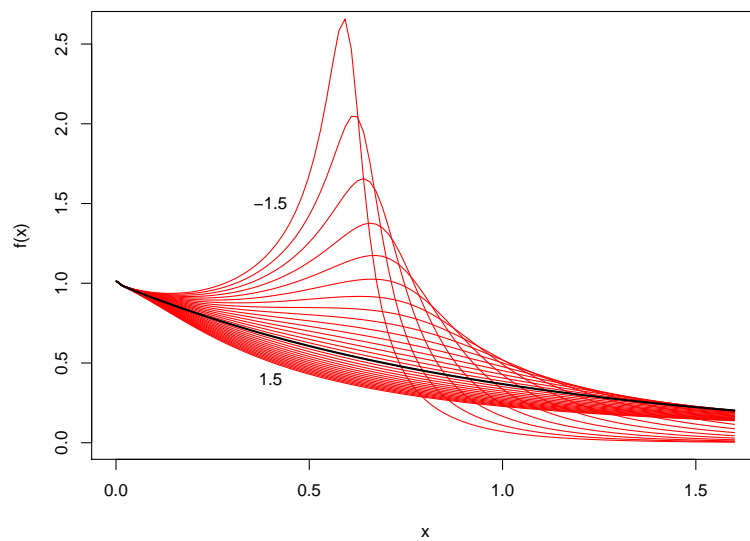
$$Q_0(u) = \frac{\log \left(\log \left(\frac{1}{1-u} \right) \right)}{k} + \log(\beta).$$

□

Tästä huomataan, että eksponentti-polynominen kvantiilisekoitus on itse asiassa Weibull-jakauman yleistys. Weibull-jakaumaan päädytään, kun asetetaan $b = \frac{1}{k}$, $a_2 = 0$, $a_1 = 0$ ja $a_0 = \log(\beta)$. Seuraavissa kuvissa 6 ja 7 on havainnollistettu, millainen vaikutus komponenteilla a_1 ja a_2 on Weibull(1,1)-jakauman eli Exp(1)-jakauman tiheysfunktioon, kun niitä poikkeutetaan nolasta.



Kuva 6: Kertoimen a_1 vaikutus Weibull(1,1)-jakauman tiheysfunktioon kun sitä poikkeutetaan nolasta. Tässä $a_1 = -1.5, -1, 4, \dots, 1.4, 1.5$, ja tapaus $a_1 = 0$ on korostettuna.



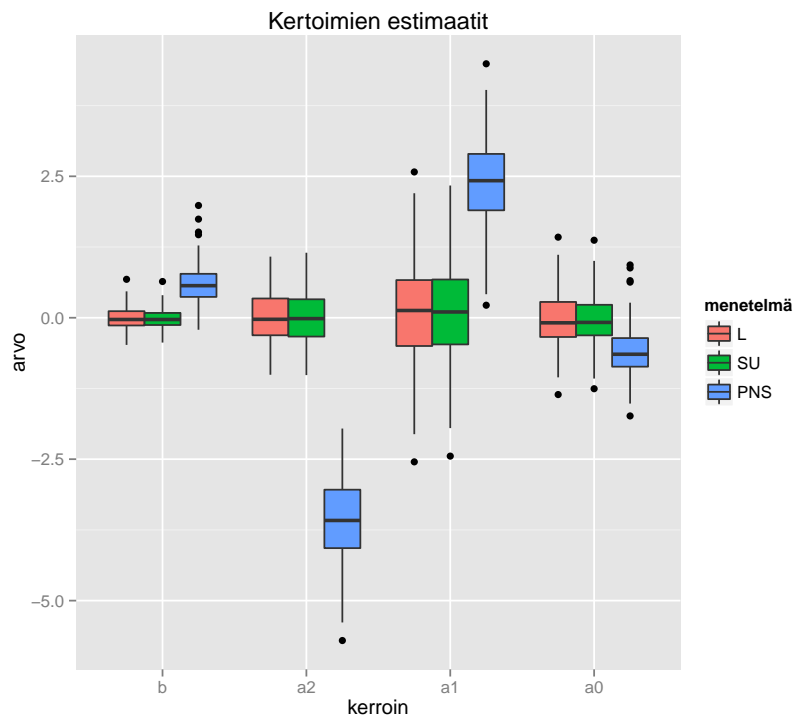
Kuva 7: Kertoimen a_2 vaikutus Weibull(1,1)-jakauman tiheysfunktioon kun sitä poikkeutetaan nolasta. Tässä $a_2 = -1.5, -1, 4, \dots, 1.4, 1.5$, ja tapaus $a_2 = 0$ on korostettuna.

4.4.2 Simulointiesimerkki

Tarkastellaan vastaavasti eksponentti-polynomisen kvantiilisekoituksen tapauksessa kolmen estimointimenetelmän käyttäytymistä simuloimalla kvantiilisekoitusjakaumasta 200 kertaa 200 havainnon otos. Käytetään jakaumaa, jonka parametrien arvot ovat $b = 1$, $a_2 = -0.5$, $a_1 = -1$ ja $a_0 = 0$. Kyseessä on siis jakauma, jonka kvantiilifunktio on

$$Q(u) = \log\left(\log\left(\frac{1}{1-u}\right)\right) - 0.5u^2 - u, \quad 0 < u < 1. \quad (27)$$

Estimoidaan kvantiilisekoituksen parametrit kolmella menetelmällä kullekin simulointitokselle ja tarkastellaan näiden parametriestimaattien jakaumaa.



Kuva 8: Eksponentti-polynomisesta kvantiilisekoitusjakaumasta poimituille otoksille eri menetelmällä laskettujen oikean arvon suhteen keskitettyjen parametriestimaattien viiksilaattikokuvaajat.

Menetelmistä SU-menetelmä ja L-momenttimenetelmä antavat lähes yhtä hyviä tuloksia, mutta PNS-menetelmällä saaduissa estimaateissa on paljon harhaa yksittäisten kertoimien kohdalla. Erityisesti harhaa on kertoimissa a_2 ja a_1 . PNS-menetelmä ei vaikuta kovin käyttökelpoiselta tämän kvantiilisekoituksen kohdalla.

Taulukko 2: Edellisistä 200 otoksesta eri menetelmillä saatujen parametrivektoreiden keskiarvot ja niitä vastaavien sovitteiden ja simulointijakauman välisten Kolmogorov – Smirnov -tunnuslukujen keskiarvot sekä kaikkien näiden keskihajonnat.

Parametri	L-momenttimenetelmä	SU-menetelmä	PNS-menetelmä
$b = 1$	0.985 (0.173)	0.980 (0.151)	1.593 (0.320)
$a_2 = -0.5$	-0.516 (0.440)	-0.508 (0.435)	-4.064 (0.730)
$a_1 = -1$	-0.907 (0.840)	-0.905 (0.819)	1.397 (0.760)
$a_0 = 0$	-0.051 (0.426)	-0.059 (0.392)	-0.600 (0.422)
K-S	0.042 (0.018)	0.039 (0.017)	0.499 (0.110)

Taulukossa 2 on estimoidun parametrivektorin keskiarvot ja keskihajonnat menetelmittäin. Huomataan, että L-momentti- ja SU-estimaatit ovat hyvin lähellä toisiaan, mutta PNS-estimaatit eroavat merkittävästi näistä kahdesta.

4.5 Gamma-polynominen kvantiilisekoitus

EkspONENTTijakauma pohjajakaumana on itse asiassa Gamma(1, 1)-jakauma. Tarkastellaan tämän vuoksi toista mahdollista gammajakaumaa kvantiilisekoituksen pohjajakaumana. Otetaan pohjajakaumaksi Gamma(1, 2). Tässä ensimmäinen parametri on käänteinen skaala- ja toinen muotoparametri. Olkoon $X \sim \text{Gamma}(1, 2)$. Silloin

$$f_X(x) = xe^{-x}, \quad x > 0. \quad (28)$$

Vastaavasti tehdään X :lle logaritmuunnos, jotta se soveltuu kvantiilisekoituksen pohjajakaumaksi. Tämän logaritmoidun muuttujan kertymäfunktio on

$$F_0(x) = F_{\log X}(x) = 1 - e^{-e^x}(1 + e^x). \quad (29)$$

Tarvittava kvantiilifunktio Q_0 voidaan hakea numeerisesti tämän käänteisfunktiona F_0^{-1} . Nimitetään tätä mallia gamma-polynomiseksi kvantiilisekoitukseksi.

Olkoon logaritmoitu otos y_1, \dots, y_n . Estimointi voidaan tehdä kolmella luvussa 3 esitetyllä tavalla.

L-momenttiestimaattia varten otoksesta voidaan laskea kaavalla (9) L-momenttien estimaatit $\bar{l}_1, \bar{l}_2, \bar{l}_3, \bar{l}_4$. Itse estimointiyhtälöryhmä (13) saa gamma-polynomisen kvantiilisekoituksen tapauksessa muodon:

$$\begin{pmatrix} \bar{l}_1 \\ \bar{l}_2 \\ \bar{l}_3 \\ \bar{l}_4 \end{pmatrix} = \begin{pmatrix} \lambda_{1,\log X} & \frac{1}{3} & \frac{1}{2} & 1 \\ \lambda_{2,\log X} & \frac{1}{6} & \frac{1}{6} & 0 \\ \lambda_{3,\log X} & \frac{1}{30} & 0 & 0 \\ \lambda_{4,\log X} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b \\ a_2 \\ a_1 \\ a_0 \end{pmatrix}, \quad (30)$$

jossa $\lambda_{i,\log(X)}$, $i = 1, \dots, 4$, saadaan kaavasta (8) käyttäen muuttujan $\log(X)$ kvantiilifunktiota Q_0 :

$$\lambda_{1,\log X} = 0.423, \quad \lambda_{2,\log X} = 0.443, \quad \lambda_{3,\log X} = -0.053 \quad \text{ja} \quad \lambda_{4,\log X} = 0.061.$$

Estimoidun jakauman tiheysfunktio saadaan vastaavasti, kuten kaavassa (20), ja se saa logaritmisen Gamma(1, 2)-pohjafunktion tilanteessa muodon:

$$f_Y(y) = f_Y(Q(u)|\tilde{b}, \tilde{a}_2, \tilde{a}_1, \tilde{a}_0) = \frac{1}{\frac{\tilde{b}}{F_0'(Q_0(u))} + 2\tilde{a}_2u + \tilde{a}_1} = \frac{1}{\frac{\tilde{b}}{e^{(2Q_0(u) - e^{Q_0(u)})}} + 2\tilde{a}_2u + \tilde{a}_1}. \quad (31)$$

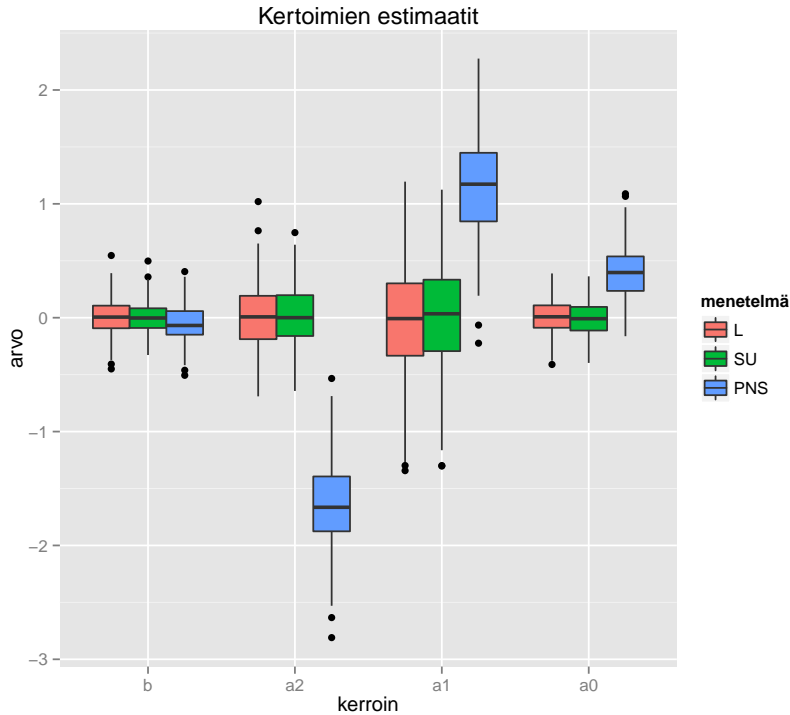
Kaavalla (18) saadaan tästä alkuperäisen skaalan tiheysfunktio. Jakauman funktiot ja parametrien estimointi L-momenttien avulla ovat R-koodissa liitteessä 3. SU- ja PNS-estimaatit saadaan edellä esitetyllä tavalla.

4.5.1 Simulointiesimerkki

Tarkastellaan vastaavasti gamma-polinomisen kvantiilisekoituksen tapauksessa kolmen estimointimenetelmän käyttäytymistä simuloimalla kvantiilisekoitusjakaumasta 200 kertaa 200 havainnon otos. Käytetään jakaumaa, jonka parametrien arvot ovat $b = 1$, $a_2 = -0.5$, $a_1 = -1$ ja $a_0 = 0$. Kyseessä on siis jakauma, jonka kvantiilifunktio on

$$Q(u) = Q_0(u) - 0.5u^2 - u, \quad 0 < u < 1. \quad (32)$$

Estimoidaan kvantiilisekoituksen parametrit kolmella menetelmällä kullekin simulointitokselle ja tarkastellaan näiden parametriestimaattien jakaumaa.



Kuva 9: Gamma-polinomisesta kvantiilisekoitusjakaumasta poimituille otoksille eri menetelmällä laskettujen oikean arvon suhteen keskitettyjen parametriestimaattien viiksilaatikkokuvaajat.

SU-menetelmällä ja L-momenttimenetelmällä saadaan lähes yhtä hyviä tuloksia. PNS-menetelmän tulokset poikkeavat näistä jälleen varsin paljon juurikin kertoimien a_2 ja a_1 osalta. Näiden kertoimien vaikutukset sekoittuvat PNS-estimaatissa ja hajaantuvat kahteen suuntaan. Vaikuttaa siltä, että PNS-menetelmä toimii korkeilla tapauksista vain normaalissa tilanteessa hyvin. Kun pohjajakaumaksi otetaan vinompia vaihtoehtoja, niin estimaatteihin tulee harhaa.

Taulukko 3: Edellisistä 200 otoksesta eri menetelmillä saatujen parametrivektoreiden keskiarvot ja niitä vastaavien sovitteiden ja simulointijakauman välisten Kolmogorov – Smirnov -tunnuslukujen keskiarvot sekä kaikkien näiden keskihajonnat.

Parametri	L-momenttimenetelmä	SU-menetelmä	PNS-menetelmä
$b = 1$	1.007 (0.165)	0.995 (0.143)	0.943 (0.171)
$a_2 = -0.5$	-0.495 (0.286)	-0.487 (0.248)	-2.154 (0.402)
$a_1 = -1$	-1.019 (0.486)	-0.994 (0.472)	0.155 (0.444)
$a_0 = 0$	0.004 (0.148)	-0.006 (0.144)	0.390 (0.231)
K-S	0.056 (0.027)	0.042 (0.019)	0.880 (0.105)

Taulukossa 3 on estimoidun parametrivektorin keskiarvot ja keskihajonnat menetelmittäin. Huomataan, että L-momentti- ja SU-estimaatit ovat hyvin lähellä toisiaan, mutta PNS-estimaatit eroavat merkittävästi näistä kahdesta.

5 Kvantiilisekoitusten sovittumistarkasteluja

Polynomisen kvantiilisekoitusmallin käytettävyyden kannalta oleellista on sen kyky sovittua erilaisiin elinaika-aineistoihin. Edellä tarkasteltiin mm. kvantiilisekoitusten parametrien estimointia, kun malli oli oikea, jolloin aineisto noudatti kvantiilisekoitusjakaumaa. Mitä tapahtuu, kun kvantiilisekoitusmalli ei ole oikea, vaan aineisto tulee jostain toisesta jakaumasta? Tarkastellaan tässä luvussa tällaista tilannetta.

5.1 Weibull-jakauman ominaisuuksista

Tarkastellaan tässä luvussa polynomisten kvantiilisekoitusten kykyä sovittua jakauksiin, jotka ovat riskifunktioltaan muutamaa perustyyppiä. Haetaan nämä jakaumat Weibull-perheestä. Tätä varten seuraavassa on tehty joitain huomioita Weibull-jakaumasta.

Weibull-jakauman riskitiheys

Weibull-jakaumaperhe on erityisesti elinaikojen mallintamiseen sopiva parametrinen jakaumaperhe. Se parametrisoituu kahdella parametrilla. Parametri $\beta > 0$ on ns. skaalaparametri, joka kuvaa jakauman arvojen absoluuttista kokoa, ja $k > 0$ on muotoparametri, joka kuvaa jakauman tyyppiä. Jakauman kertymä- ja tiheysfunktiot ovat seuraavat, kun $X_{\beta,k} \sim \text{Weibull}(\beta, k)$: (Weibull, 1951)

$$F_{X_{\beta,k}}(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^k}, & \text{kun } x \geq 0 \\ 0, & \text{kun } x < 0 \end{cases} \quad (33)$$

$$f_{X_{\beta,k}}(x) = \begin{cases} \left(\frac{k}{\beta}\right)\left(\frac{x}{\beta}\right)^{k-1}e^{-\left(\frac{x}{\beta}\right)^k}, & \text{kun } x \geq 0 \\ 0, & \text{kun } x < 0. \end{cases} \quad (34)$$

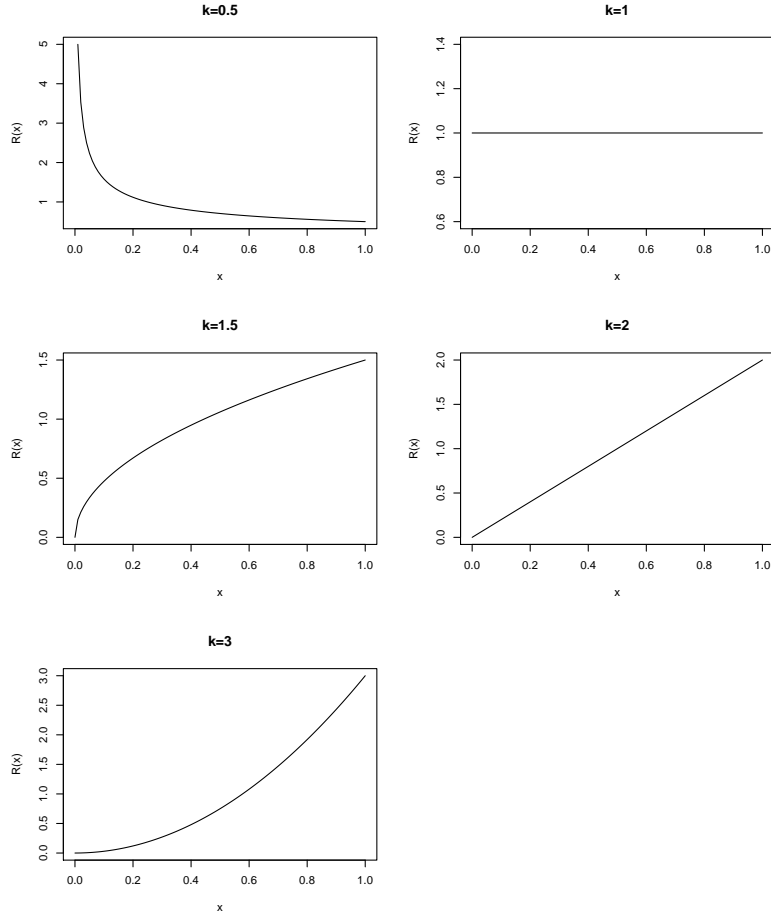
Täten jakauman riskitiheysfunktio saadaan kaavalla (16): (McCool, 2012, sivut 95 – 118)

$$R_{X_{\beta,k}}(x) = \left(\frac{k}{\beta}\right) \left(\frac{x}{\beta}\right)^{k-1}, \quad \text{kun } x \geq 0. \quad (35)$$

Tämä riskitiheysfunktio on

$$\begin{cases} \text{vähenevä,} & \text{kun } k < 1 \\ \text{vakio,} & \text{kun } k = 1 \\ \text{hidastuvasti kasvava,} & \text{kun } 1 < k < 2 \\ \text{tasaisesti kasvava,} & \text{kun } k = 2 \\ \text{kiihtyvästi kasvava,} & \text{kun } k > 2. \end{cases} \quad (36)$$

Kuvassa 10 on erilaisten Weibull-jakaumien riskitiheysfunktioita. Näissä riskitiheyksissä parametri $\beta = 1$.



Kuva 10: Erilaisten Weibull-jakaumien riskitiheysfunktioita.

Huomio logaritmisen Weibull-jakauman L-momenteista

Tarkasteluissa on tarve käyttää alkuperäiseltä skaalalta logaritmoitua muuttujaa. Sovitus voidaan tehdä tämän skaalan L-momenteja hyväksi käyttäen. Tarkastellaan siksi logaritmoidun Weibull-jakauman L-momenteja. Olkoon $X_{\beta,k} \sim \text{Weibull}(\beta, k)$. Tällöin muuttujalle $\log X_{\beta,k}$ saadaan kertymäfunktio

$$F_{\log X_{\beta,k}}(x) = F_{X_{\beta,k}}(e^x) = 1 - e^{-\left(\frac{e^x}{\beta}\right)^k}. \quad (37)$$

Tällöin tämän käänteisfunktiona saadaan sen kvantiilifunktio

$$Q_{\log X_{\beta,k}}(u) = \frac{\log(-\log(1-u))}{k} + \log(\beta), \quad \text{kun } 0 < u < 1, \quad (38)$$

kuten kaavan (26) johdossa.

Parametri β siis vaikuttaa kvantiilifunktion arvoihin vain lisäyksen $\log(\beta)$ verran. Tästä seuraa kaavalla (8), että logaritmisen Weibull-jakauman L-momenteille

$$\begin{aligned} \lambda_{1,\log X_{\beta,k}} &= \lambda_{1,\log X_{\beta=1,k}} + \log(\beta) \quad \text{ja} \\ \lambda_{i,\log X_{\beta,k}} &= \lambda_{i,\log X_{\beta=1,k}}, \quad \text{kun } i \geq 2. \end{aligned} \quad (39)$$

Perustelu.

$$\begin{aligned}
\lambda_{1,\log X_{\beta,k}} &= \int_0^1 \frac{\log(-\log(1-u))}{k} + \log(\beta) du \\
&= \int_0^1 \frac{\log(-\log(1-u))}{k} du + \int_0^1 \log(\beta) du = \lambda_{1,\log X_{\beta=1,k}} + \log \beta, \\
\lambda_{2,\log X_{\beta,k}} &= \int_0^1 (2u-1) \left[\frac{\log(-\log(1-u))}{k} + \log(\beta) \right] du \\
&= \int_0^1 (2u-1) \frac{\log(-\log(1-u))}{k} du + \int_0^1 (2u-1) \log(\beta) du \\
&= \lambda_{2,\log X_{\beta=1,k}} + \log(\beta) \int_0^1 (2u-1) du = \lambda_{2,\log X_{\beta=1,k}}.
\end{aligned}$$

Vastaavasti muille L-momenteille tämä seuraa siitä, että

$$\begin{aligned}
\int_0^1 (6u^2 - 6u + 1) du &= 0, \\
\int_0^1 (20u^3 - 30u^2 + 12u - 1) du &= 0 \text{ jne.}
\end{aligned}$$

□

Tämä tarkoittaa, että parametri β vaikuttaa vain ensimmäiseen logaritmisesta Weibull-jakauman L-momenttiin.

5.2 Kvantiilisekoitusten sovittuminen Weibull-aineistoihin

Edellä nähtiin, että Weibull-jakaumaperheeseen kuuluu laaja kirjo erityisesti riskitehtäjäntään erilaisia jakaumia. Tarkastellaan tässä näistä muutamaa perustyyppiä yksi kerrallaan. Tarkastellaan kolmen kvantiilisekoituksen kykyä sovittua näihin Weibull-jakaumiin L-momenttimenetelmällä. Kysymys on siis siitä, että miten lähelle oikeaa mallia päästään, kun mallinnetaan Weibull-aineistoa väärällä kvantiilisekoitusmallilla, jolla on samat L-momentit kuin logaritmisella Weibull-jakaumalla.

Edellä huomattiin, että logaritmisesta Weibull-jakauman parametri β vaikuttaa vain jakauman ensimmäiseen L-momenttiin. Kun L-momenteja käytetään kvantiilisekoitusmallin parametrien estimointiin, kuten yhtälöissä (14), niin ensimmäinen L-momentti $\lambda_{1,\log X_{\beta,k}}$ vaikuttaa vain parametrin a_0 estimaattiin. Siten parametrin β vaikutus periytyy vain sijaintiparametriin a_0 , kun estimoidaan kvantiilisekoitus, jolla on samat L-momentit.

Tällöin, jos R_{e^X} on logaritmiskaalan muuttujan X eksponenttimuunnoksen riskitehtäjänsfunktio, niin vastaavalla riskitehtäjänsfunktioilla muuttujalle $X - \log(\alpha)$, $\alpha > 0$, on yhteys siihen kaavalla:

$$R_{e^{X-\log(\alpha)}}(x) = \alpha R_{e^X}(\alpha x), \quad x > 0. \quad (40)$$

Perustelu. Olkoon X satunnaismuuttuja ja sen kertymäfunktio F_X ja tiheysfunktio f_X . Esitetään muuttujan e^X riskifunktio näiden avulla:

$$\begin{aligned} F_{e^X}(x) &= F_X(\log(x)), \text{ ja siten} \\ f_{e^X}(x) &= \frac{d}{dx} F_X(\log(x)) = \frac{1}{x} f_X(\log(x)), \text{ joten} \\ R_{e^X}(x) &= \frac{\frac{1}{x} f_X(\log(x))}{1 - F_X(\log(x))}. \end{aligned}$$

Esitetään sitten muuttujan $e^{X-\log(\alpha)}$ riskifunktio tämän avulla:

$$\begin{aligned} F_{e^{X-\log(\alpha)}}(x) &= P(e^{X-\log(\alpha)} \leq x) \\ &= P(X \leq \log(\alpha x)) = F_X(\log(\alpha x)) \text{ ja} \\ f_{e^{X-\log(\alpha)}}(x) &= \frac{d}{dx} F_{e^{X-\log(\alpha)}}(x) = \frac{d}{dx} F_X(\log(\alpha x)) \\ &= \frac{1}{x} f_X(\log(\alpha x)), \text{ joten} \\ R_{e^{X-\log(\alpha)}}(x) &= \frac{\frac{1}{x} f_X(\log(\alpha x))}{1 - F_X(\log(\alpha x))} = \alpha \cdot \frac{\frac{1}{\alpha x} f_X(\log(\alpha x))}{1 - F_X(\log(\alpha x))} \\ &= \alpha R_{e^X}(\alpha x). \end{aligned}$$

□

Tämän tarkastelun tarkoitus oli näyttää, että kvantiilisekoitukset soveltuvat riskitiheysfunktion mielessä tietyllä tavalla samoin Weibull-jakaumaan parametrilla k riippumatta parametrin β arvosta. Kaavassa oleva $-\log(\alpha)$ edustaa tässä edellä havaittua Weibull-jakauman parametrin β vaikutusta kvantiilisekoitussovituksen sijainnitiiparametriin a_0 . Nähtiin siis, että parametri β ei vaikuta sovitteeseen muuten kuin sijainnin kautta. Voidaan siis kiinnittää jatkossa $\beta = 1$ ilman, että tulokset muuttuisivat.

Kvantiilisekoitus, jolla on samat L-momentit, sovitetaan ratkaisemalla yhtälöryhmä

$$\begin{pmatrix} \lambda_{1,\log X_{\beta=1,k}} \\ \lambda_{2,\log X_{\beta=1,k}} \\ \lambda_{3,\log X_{\beta=1,k}} \\ \lambda_{4,\log X_{\beta=1,k}} \end{pmatrix} = \begin{pmatrix} \lambda_1(Q_0) & \frac{1}{3} & \frac{1}{2} & 1 \\ \lambda_2(Q_0) & \frac{1}{6} & \frac{1}{6} & 0 \\ \lambda_3(Q_0) & \frac{1}{30} & 0 & 0 \\ \lambda_4(Q_0) & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b \\ a_2 \\ a_1 \\ a_0 \end{pmatrix}, \quad (41)$$

jossa vasemman puolen arvot ovat logaritmisin Weibull-jakauman L-momentteja parametrin arvoilla $\beta = 1$ ja k . Kaavasta (38) itse asiassa nähdään, että $\lambda_{i,\log X_{\beta=1,k}} = \frac{1}{k} \lambda_{i,\log X_{\beta=1,k=1}}$. Tästä seuraa, että parametrin estimaatti käyttäytyy vastaavasti:

$$(\tilde{b}, \tilde{a}_2, \tilde{a}_1, \tilde{a}_0)_{\beta=1,k}^T = \frac{1}{k} (\tilde{b}, \tilde{a}_2, \tilde{a}_1, \tilde{a}_0)_{\beta=1,k=1}^T. \quad (42)$$

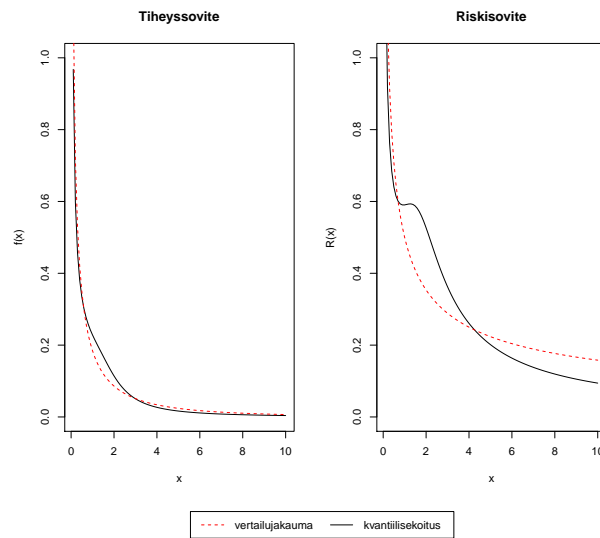
Tämä käyttäytyminen johtuu siis pelkästään valinnasta $\beta = 1$.

EkspONENTTI-polinomiselle kvantiilisekoitukselle sovite on aina täsmälleen sama kuin vastaava Weibull-jakauma, kuten nähtiin logaritmisin Weibull-jakauman kvantiilifunktiosta (26).

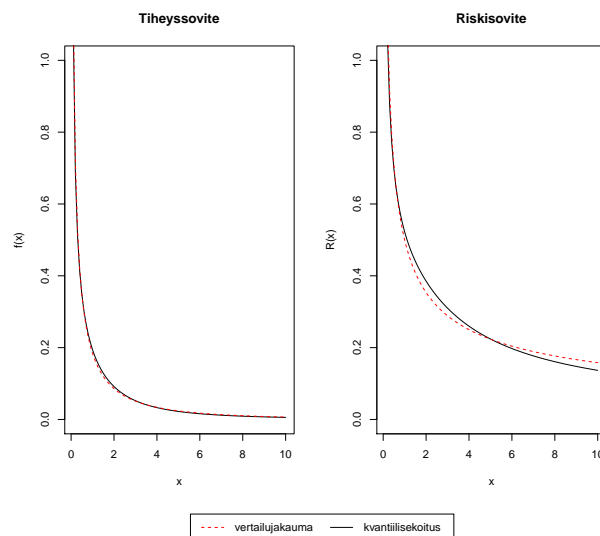
Verrataan seuraavaksi sovitettujen jakaumien tiheys- ja riskitiheysfunktioita vastaaviin Weibull-jakauman funktioihin erilaisten riskien tapauksissa. Tässä on oleellisesti kyse siitä, että kuinka hyvin kvantiilisekoitus pystyy korvaamaan Weibull-jakauman, kun halutaan käyttää riskitiheysfunktioita kuvailevana menetelmänä, ja kuinka hyvin se tekee sen erilaisten riskien tapauksessa.

Vähenevän riskin tapaus

Edellä nähtiin, että Weibull-jakauman riskitiheysfunktio on vähenevä parametrien arvoilla $0 < k < 1$. Tarkastellaan tässä erityisesti tilannetta $k = 0.5$. Kahdella eri L-momenttisovituksella saadaan seuraavat sovitteet kuvissa 11 ja 12, ja taulukossa 4 on parametrien estimaatit ja tunnuslukuja.



Kuva 11: Lognormaali-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktioita ja vastaavat Weibull($1, \frac{1}{2}$)-jakaumalle.



Kuva 12: Gamma-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktioita ja vastaavat Weibull($1, \frac{1}{2}$)-jakaumalle.

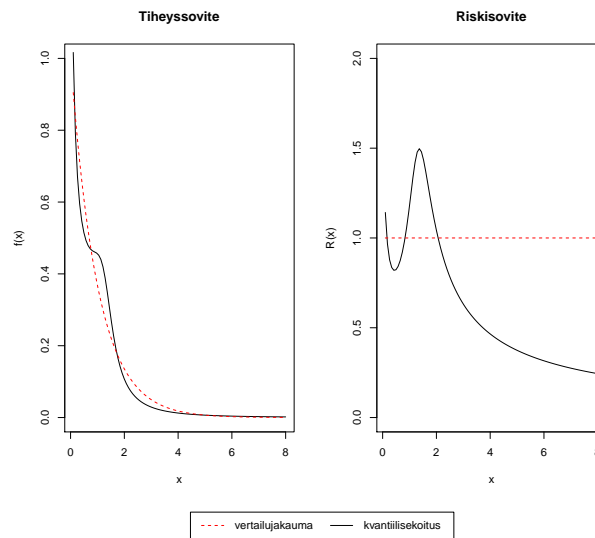
Taulukko 4: Eri kvantiilisekoitusten parametreja tapauksessa $k=0.5$.

Parametri	Lognorm.	Eksp.	Gamma
\tilde{b}	3.014	2	3.400
\tilde{a}_2	-7.067	0	-1.664
\tilde{a}_1	5.183	0	0.938
\tilde{a}_0	-1.390	0	-2.506

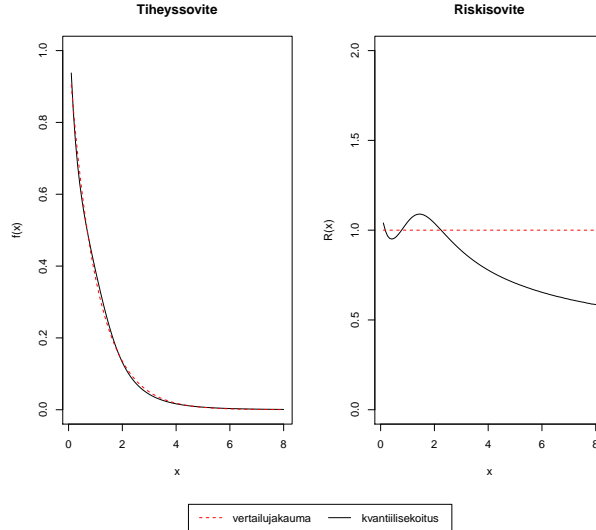
Kaikki tarkastelussa olleet kvantiilisekoitukset soveltuvat varsin hyvin tällaisen riskin mallintamiseen. Normaalikomponentti yhdistettynä polynomiin kuitenkin tuottaa tietynlaisen oskillaatioilmiön. Gamma-polynominen on varsin lähellä kohdejakaumaa. Kvantiilisekoituksen soveltuvat riskiltään tällaisen jakauman mallintamiseen hyvin.

Vakioriskin tapaus

Arvolla $k = 1$ tavoiteltavana Weibull-jakaumana on eksponenttijakauma. Tällöin kohteen riskitiheysfunktio on vakio. Kahdella eri L-momenttisovituksella saadaan seuraavat sovitteet kuvissa 13 ja 14, ja taulukossa 5 on parametrien estimaatit ja tunnuslukuja.



Kuva 13: Lognormaali-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktio ja vastaavat Weibull(1, 1)-jakaumalle.



Kuva 14: Gamma-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktio ja vastaavat Weibull(1, 1)-jakaumalle.

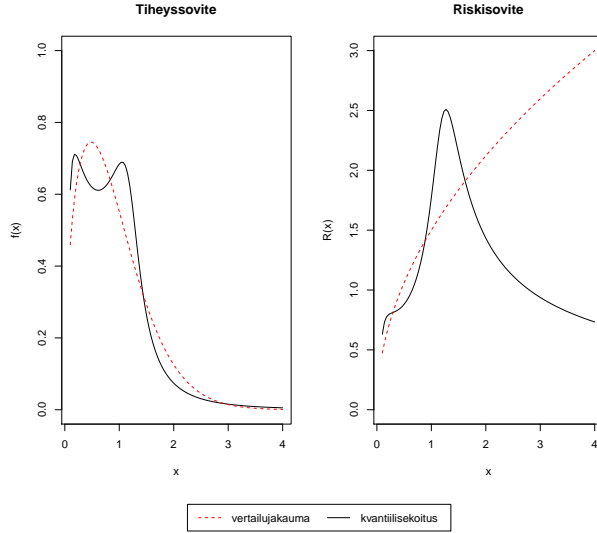
Taulukko 5: Eri kvantiilisekoitusten parametreja tapauksessa $k=1$.

Parametri	Lognorm.	Eksp.	Gamma
\tilde{b}	1.507	1	1.700
\tilde{a}_2	-3.534	0	-0.832
\tilde{a}_1	2.592	0	0.469
\tilde{a}_0	-0.695	0	-1.253

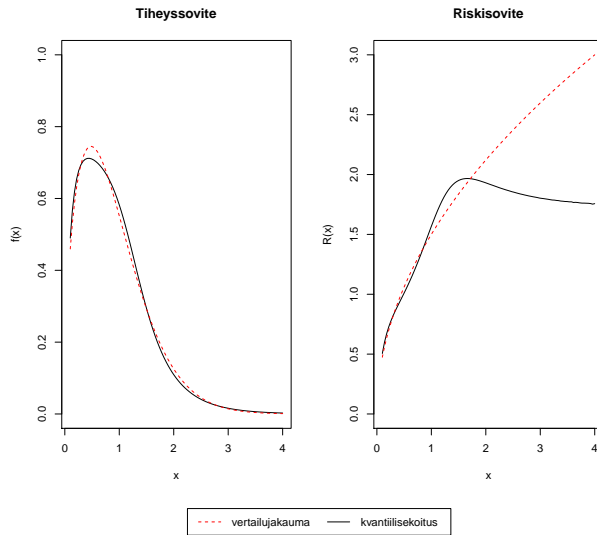
Vakioriskin tapaus on kvantiilisekoituksille sinänsä jakauman sovittavuuden kannalta aivan toimiva. Lognormaali- ja gamma-polynomiset mallit eivät kuitenkaan pysty kuvaamaan kovinkaan hyvin itse riskifunktion käyttäytymistä. Riskifunktioissa havaitaan voimakas oskillaatioilmiö näissä kahdessa tapauksessa. Jos riskitiheysfunktio on kiinnostuksen kohteena, niin silloin näiden kahden käyttö ei ole kannattavaa. Eksponentti-polynomisen kvantiilisekoitus toimii luonnollisesti tässäkin tapauksessa.

Hidastuvasti kasvavan riskin tapaus

Arvolla $1 < k < 2$ Weibull-jakauman riskitiheysfunktio on hidastuvasti kasvava. Poimitaan näistä tapaus $k = \frac{3}{2}$. Kahdella eri L-momenttisovituksella saadaan seuraavat sovitteet kuvissa 15 ja 16, ja taulukossa 6 on parametrien estimaatit ja tunnuslukuja.



Kuva 15: Lognormaali-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktio ja vastaavat Weibull($1, \frac{3}{2}$)-jakaumalle.



Kuva 16: Gamma-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktio ja vastaavat Weibull($1, \frac{3}{2}$)-jakaumalle.

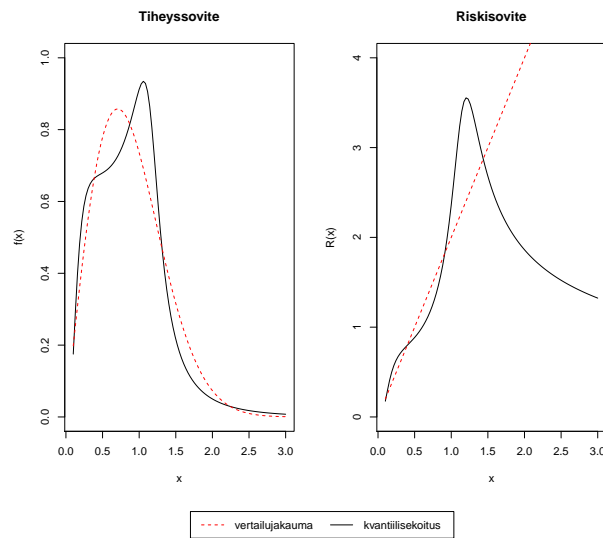
Taulukko 6: Eri kvantiilisekoitusten parametreja tapauksessa $k = \frac{3}{2}$.

Parametri	Lognorm.	Eksp.	Gamma
\tilde{b}	1.005	$\frac{2}{3}$	1.133
\tilde{a}_2	-2.356	0	-0.555
\tilde{a}_1	1.728	0	0.313
\tilde{a}_0	-0.463	0	-0.835

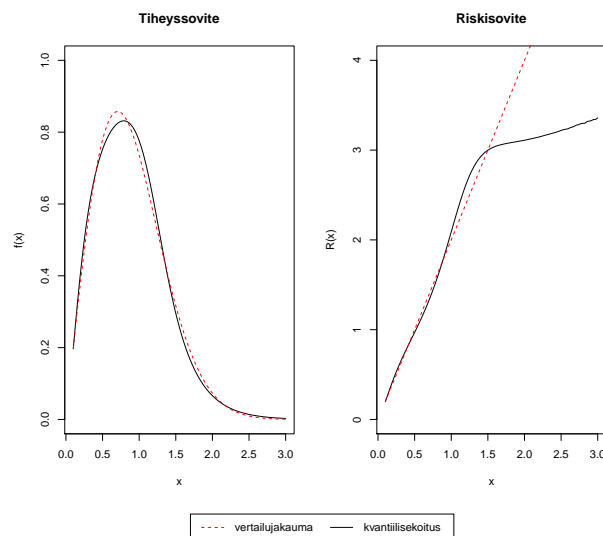
Hidastuvasti kasvavassa tapauksessa gamma-polynominen kvantiilisekoitus sovituu varsin hyvin. Se löytää myös riskifunktion tyypin kohtuullisesti, mutta käyttäytyminen jakauman hännässä ei vastaa kovin hyvin kohdejakaumaa. Lognormaali-polynominen antaa heikohkon tuloksen, koska se ei löydä jakauman yksihuippuisuutta. Sen riskitiheysfunktio sovituu myös melko heikosti, jos verrataan gamma-polynomiseen kvantiilisekoitukseen.

Tasaisesti kasvavan riskin tapaus

Arvolla $k = 2$ Weibull-jakauman riskitiheysfunktio on tasaisesti kasvava. Kahdella eri L-momenttisovituksella saadaan seuraavat sovitteet kuvissa 17 ja 18, ja taulukossa 7 on parametrien estimaatit ja tunnuslukuja.



Kuva 17: Lognormaali-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktio ja vastaavat Weibull(1, 2)-jakaumalle.



Kuva 18: Gamma-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktio ja vastaavat Weibull(1, 2)-jakaumalle.

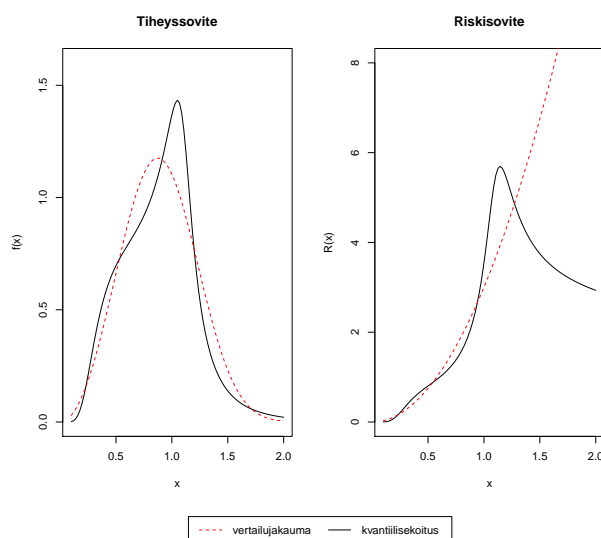
Taulukko 7: Eri kvantiilisekoitusten parametreja tapauksessa $k = 2$.

Parametri	Lognorm.	Eksp.	Gamma
\tilde{b}	0.753	$\frac{1}{2}$	0.850
\tilde{a}_2	-1,767	0	-0.416
\tilde{a}_1	1.296	0	0.234
\tilde{a}_0	-0.348	0	-0.626

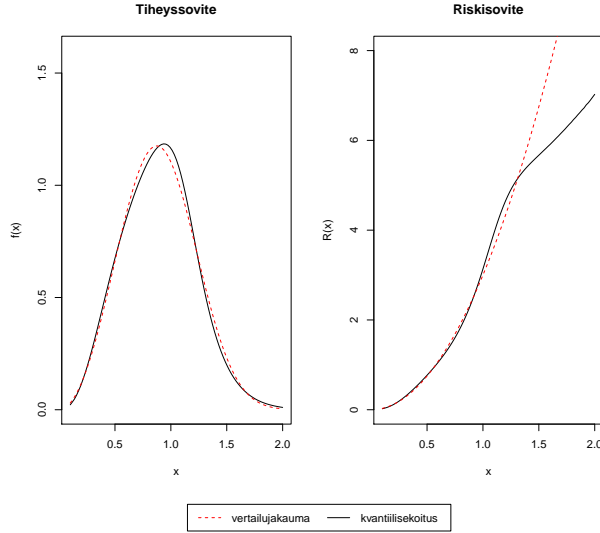
Tässä tilanteessa lognormaali-polynomisen sekoitus löytää itse asiassa jakauman varsin hyvin. Myös riskitiheysfunktio kuvaa hyvin todellisuutta oleellisella alueella. Jakauman häntä on kuitenkin varsin erilainen kuin kohdejakaumalla. Gamma-polynomisen kvantiilisekoitus sovituu taas selvästi paremmin. Sen riskitiheysfunktion sovite on myös varsin hyvä.

Kiihtyvästi kasvavan riskin tapaus

Arvolla $k > 2$ Weibull-jakauman riskitiheysfunktio on kiihtyvästi kasvava. Tarkastellaan tapausta $k = 3$. Kahdella eri L-momenttisovituksella saadaan seuraavat sovitteet kuvissa 19 ja 20, ja taulukossa 8 on parametrien estimaatit ja tunnuslukuja.



Kuva 19: Lognormaali-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktio ja vastaavat Weibull(1, 3)-jakaumalle.



Kuva 20: Gamma-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktioit ja vastaavat Weibull(1, 3)-jakaumalle.

Taulukko 8: Eri kvantiilisekoitusten parametreja tapauksessa $k = 3$.

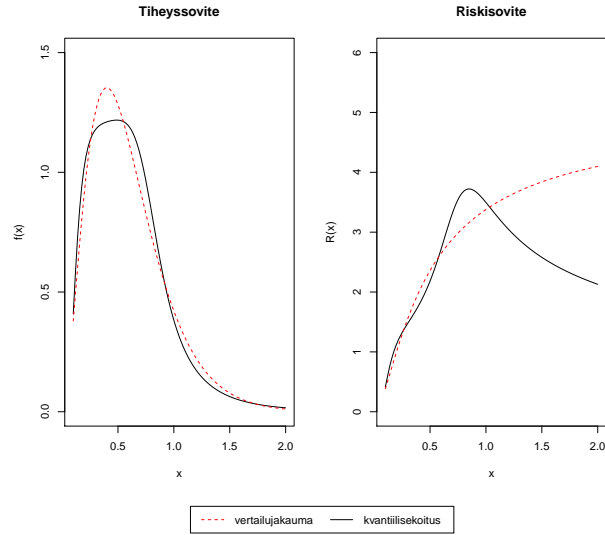
Parametri	Lognorm.	Eksp.	Gamma
\tilde{b}	0.502	$\frac{1}{3}$	0.567
\tilde{a}_2	-1,778	0	-0.277
\tilde{a}_1	0.864	0	0.156
\tilde{a}_0	-0.232	0	-0.418

Kiihtyvästi kasvavan riskin tapaus on sinänsä samanlainen kuin edeltävä, mutta sovittavuus tässä tapauksessa on edelleen parempi kuin edellisessä kummankin jakauman kohdalla. Erityisesti gamma-polynomisen kvantiilisekoitus on myös riskifunktioltaan lähellä oikeaa.

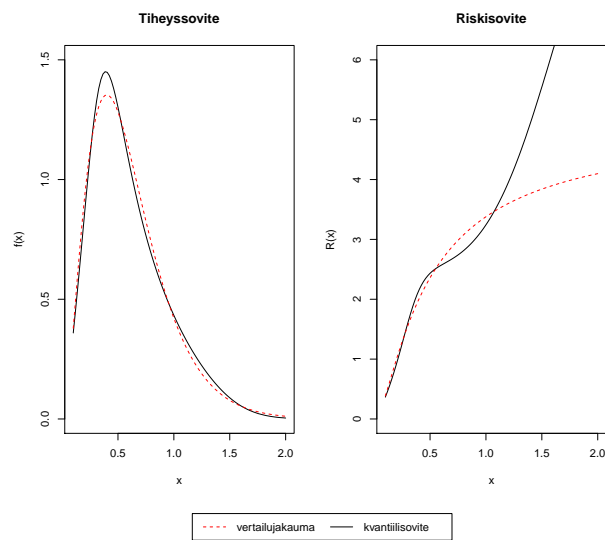
Kaiken kaikkiaan voidaan todeta, että Weibull-mallin korvaaminen kvantiilisekoitusmallilla voi olla harkittavissa oleva vaihtoehto. Erityisesti eksponenttipolynomisen kvantiilisekoitus on juuri Weibull-tapauksissa toimiva, sama kuin Weibull-jakauma. Gamma-polynomisen kvantiilisekoitus on myös hyvin lähellä edellistä. Lognormaali-polynomisella sekoituksella on selvät heikkoutensa näihin kahteen verrattuna. On kuitenkin huomattava, että tässä on rajoitettu vain Weibull-jakaumaan.

Vaihtoehto gammajakaumasta

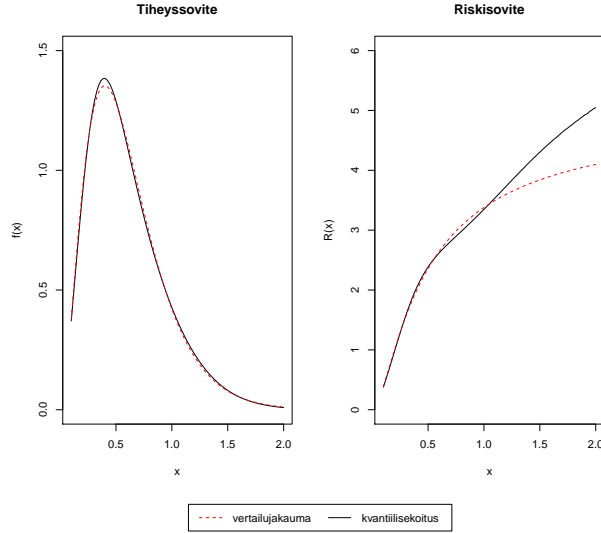
Weibull-jakaumien lisäksi tarkastellaan sovittumista Gamma(3, 5)-jakaumaan. Parametrit ovat käänteinen skaala- ja muotoparametri. Jakauma muistuttaa monin tavoin Weibull-jakaumaa parametrin arvolla $k = \frac{3}{2}$, riskitiheys on hidastuvasti kasvava. Kolmella eri L-momenttisovituksella saadaan seuraavat sovitteet kuvissa 21, 22 ja 23, ja taulukossa 9 on parametrien estimaatit ja tunnuslukuja.



Kuva 21: Lognormaali-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktiot ja vastaavat Gamma(3, 5)-jakaumalle.



Kuva 22: Eksponentti-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktiot ja vastaavat Gamma(3, 5)-jakaumalle.



Kuva 23: Gamma-polynomisen kvantiilisekoitussovitteen tiheys- ja riskitiheysfunktioit ja vastaavat Gamma(3, 5)-jakaumalle.

Taulukko 9: Eri kvantiilisekoitusten parametreja tapauksessa Gamma(3, 5).

Parametri	Lognorm.	Eksp.	Gamma
\tilde{b}	0.672	0.446	0.758
\tilde{a}_2	-1.017	0.559	0.188
\tilde{a}_1	0.839	-0.317	-0.107
\tilde{a}_0	-0.767	-0.457	-1.016

Sovitteet ovat jokseenkin samanlaisia kuin aikaisemmin Weibull($1, \frac{3}{2}$)-jakaumalle saadut. Lognormaali-polynomisen sovittuu jälleen heikommin kuin kaksi muuta. Hieman yllättäen gamma-polynomisen kvantiilisekoitus sovittuu nyt jopa hieman paremmin kuin eksponentti-polynomisen versio. Eksponentti-polynomisen kvantiilisekoitus ei siis välttämättä ole ylivertainen gamma-polynomiseen verrattuna, kuten Weibull-esimerkeistä voisi olettaa.

Yleisesti voidaan sanoa, että kvantiilisekoitusmalli soveltuu tässä esitettyjen elin-aikamallien korvaajaksi melko vaihtelevasti. Jos riskifunktion tyyppi on hyvin selkeästi kasvava tai vähenevä, niin lopputulos on hyvä. Muussa tapauksessa malli voi antaa harhaanjohtavia tuloksia.

Sovitteista voi kuitenkin todeta, että kvantiilisekoitukset voivat olla harkittavissa oleva malli tilanteisiin, joissa riskifunktiossa tapahtuu jollain muuttujan arvolla muutos monotonisuuden suunnassa. Weibull-jakauma sellaisenaan ei sisällä näitä vaihtoehtoja.

6 Sovellus matkapuhelinaineistoon

Sovelletaan kvantiilisekoitusmalleja aineistoon (Karvanen ja Luoma, 2014), joka liittyy ihmisten matkapuhelimen käyttöelinkaarien pituuksien ja merkkiuskollisuuden mallintamiseen liittyvään tutkimukseen (Karvanen et al., 2014).

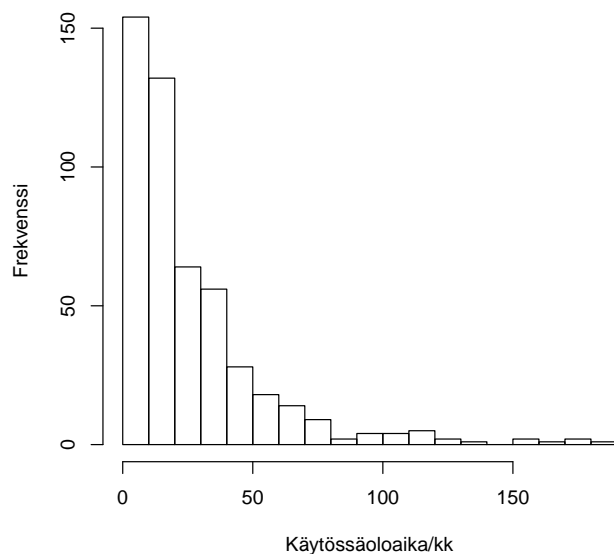
6.1 Aineisto

Tutkimusta varten on tehty kysely eri-ikäisille ja eri puolilla Suomea asuville ihmisille. Näiltä on kysytty mm. tällä hetkellä käytössä olevan puhelimen hankkimisesta kulunut aika. Vastaus on voitu antaa erilaisilla tarkkuuksilla: kuukauden, vuodenajan tai vuoden tarkkuudella. Vastemuuttujan arvo on siis välisensuroitu eli sen tiedetään olevan vain jollakin tunnetulla välillä. Havainnot ovat siis muotoa $x_i \in (c_l^i, c_u^i)$, missä c_l^i ja c_u^i määräytyvät henkilön antaman ajankohdan ja tarkkuuden mukaan. Taustamuuttujina kyselyn henkilöiltä on kirjattu mm. asuinalue, ikä, tulot, sukupuoli ja puhelimen tyyppi. Lopullisesti kyselyyn on saatu 536 vastausta. Kun karsitaan dataa niin, että ne havainnot, jotka eivät sisällä informaatiota kulu-neesta ajasta, jätetään pois, niin lopulliseen aineistoon jää 499 havaintoa. Aineisto, jota käytetään, on esitetty liitteessä 2. Kun yhdistetään ikä-muuttujan luokkia niin, että muodostuu kolme erilaista ikäryhmää: 15 – 34, 35 – 54 ja 55 – 79 -vuotiaat, niin aineisto on jakautunut iän ja asuinalueen mukaan taulukossa 10 esitettävällä tavalla.

Taulukko 10: Otoksen frekvenssit iän ja asuinalueen mukaan.

Alue/Ikä	15 – 34	35 – 54	55 – 79	Yhteensä
Uusimaa	27	42	66	135
Muu Etelä-Suomi	25	23	35	83
Länsi-Suomi	51	41	56	148
Pohjois- ja Itä-Suomi	41	45	47	133
Yhteensä	144	151	204	499

Kun poimitaan vastevälien (c_l^i, c_u^i) keskipisteet, niin koko dataa voidaan havainnollistaa histogrammilla kuvassa 24. Aineisto on tyypillinen elinaika-aineisto, jossa suurin osa puhelimista on varsin uusia, ja vanhempia puhelimia on vähän, mutta ne voivat olla iältään huomattavan vanhoja.



Kuva 24: Aineiston vastevälien keskipisteiden histogrammi.

6.2 Tutkimuskysymys

Lähdetään aineiston avulla tekemään päätelmiä matkapuhelinten iän mediaanista aineiston eri ositteissa. Pyritään vastamaan kysymykseen, että eroavatko mediaanit eri ositteissa. Jaetaan aineisto osiin kahden muuttujan suhteen ja tarkastellaan näitä osituksia toisistaan erillään. Tutkitaan, että onko puhelinten iän mediaaneissa eroa em. kolmessa ikäryhmässä, ja toiseksi tutkitaan, että poikkeako länsisuomalaisten vastaava mediaani muun maan mediaanista. Muulla Suomella tarkoitetaan tässä Uudenmaan, Muun Etelä-Suomen ja Pohjois- ja Itä-Suomen yhdistettyä otosta.

6.3 Mallintaminen

Koska tarkkoja ikä ei tiedetä, vaan puhelimille on tiedossa vain väli, jolle ikä sijoittuu, niin varsinaisia otosmediaaneja ei ole mahdollista laskea. Jotta mediaaneille saataisiin estimaatit, niin aineistolle on asetettava esimerkiksi parametrinen malli. Mallinnetaan aineistoa em. kvantiilisekoitusmallien avulla. Tätä varten tehdään aineiston vasteväleille logaritminmuunnos. Estimointi voidaan tehdä SU-menetelmällä välisensuroidulle datalle muodostamalla uskottavuusfunktio, kuten kaavassa 6.

Mediaanin tutkiminen kvantiilisekoitusten avulla onnistuu helposti, sillä mediaanille Md pätee

$$\log(Md) = Q\left(\frac{1}{2}\right) = bQ_0\left(\frac{1}{2}\right) + a_2\left(\frac{1}{2}\right)^2 + a_1\left(\frac{1}{2}\right) + a_0, \quad (43)$$

joten sen estimaatti saadaan helposti kertoimien estimaattien avulla. Tästä saadaan mediaanin estimaatti $\hat{M}d = e^{\log(\hat{M}d)}$ alkuperäisellä skaalalla. Estimaatin tarkkuutta voi arvioida esimerkiksi laskemalla mediaanille varianssiestimaatti jackknifemenetelmän avulla.

6.3.1 Ikäryhmät

Sovitetaan kvantiilisekoitusmalli erikseen kunkin ikäryhmän muodostamaan osadaan. Toistetaan tämä kullekin kolmelle kvantiilisekoitustyyppille. Verrataan näitä sovitteita eri ikäryhmien välillä.

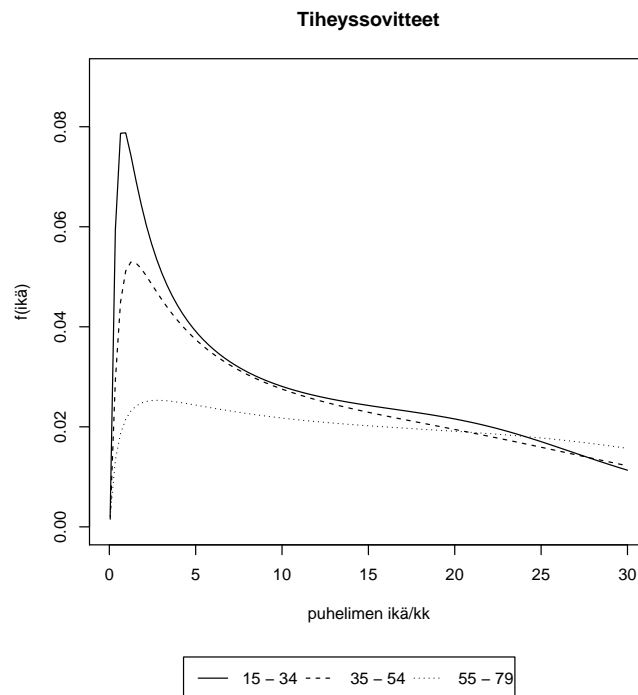
Lognormaali-polynomisen kvantiilisekoitus

Taulukossa 11 on lognormaali-polynomisen kvantiilisekoituksen sovitteet ikäryhmittäin, mediaanin estimaatti ja sen varianssin jackknife-estimaatti.

Taulukko 11: Lognormaali-polynomisen kvantiilisekoituksen sovitteet ikäryhmittäin.

Estimaatti	15 – 34	35 – 54	55 – 79
\hat{b}	1.147	1.313	1.707
\hat{a}_2	-3.151	-2.192	-1.671
\hat{a}_1	3.397	1.812	-0.287
\hat{a}_0	1.605	2.350	3.726
$\hat{M}d$	12.374	14.996	23.696
$\hat{V}ar_{jack}(Md)$	1.714	2.127	2.652

Kuvassa 25 on piirrettynä sovitteiden tiheysfunktiot.



Kuva 25: Lognormaali-polynomisen kvantiilisekoituksen sovitteiden tiheysfunktiot kolmessa eri ikäryhmässä.

Eri ryhmien mediaanien erotuksille saadaan estimaatit, keskivirheet ja Bonferroni-korjatut p-arvot taulukossa 12. Näyttää siltä, että vanhimmalla ikäryhmällä käytössä olevat puhelimet ovat mediaani-ikänsä vanhimpia. Nuoremmilla ikäryhmillä ei havaita merkitsevää eroa.

Taulukko 12: Lognormaali-polynomisen kvantiilisekoituksen ikäryhmien mediaanien vertailu.

Erotus	Arvo	sd	p-arvo
$\hat{M}d_{35-54} - \hat{M}d_{15-34}$	2.621	1.960	0.54
$\hat{M}d_{55-79} - \hat{M}d_{15-34}$	11.322	2.090	< 0.001
$\hat{M}d_{55-79} - \hat{M}d_{35-54}$	8.700	2.186	< 0.001

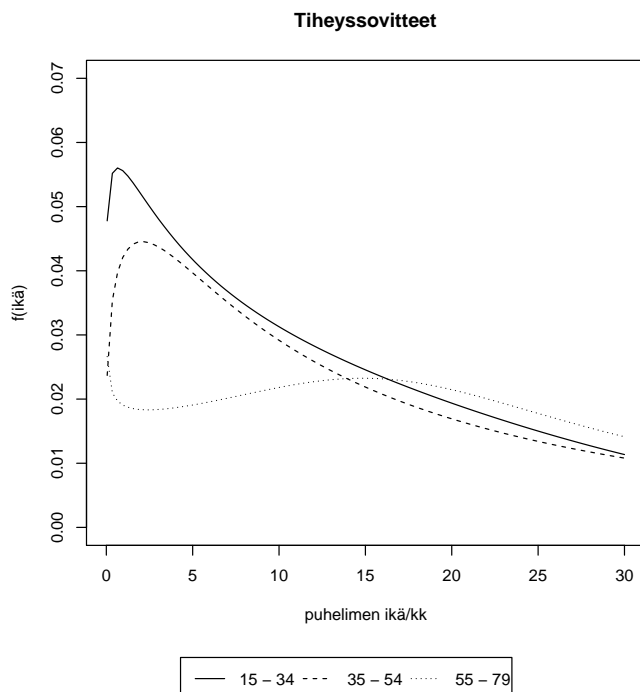
EkspONENTTI-POLYNOMINEN KVANTIILISEKOITUS

Taulukossa 13 on vastaavat eksponentti-polynomisen kvantiilisekoituksen sovitteet ikäryhmittäin, mediaanin estimaatti ja sen varianssin jackknife-estimaatti.

Taulukko 13: EkspONENTTI-POLYNOMINEN KVANTIILISEKOITUKSEN SOVITTEET IKÄRYHMITÄIN.

Estimaatti	15 – 34	35 – 54	55 – 79
\hat{b}	0.915	0.813	1.157
\hat{a}_2	-0.520	0.131	1.442
\hat{a}_1	0.780	0.579	-2.533
\hat{a}_0	2.597	2.669	4.498
$\hat{M}d$	12.446	14.771	23.747
$\hat{V}ar_{jack}(Md)$	4.427	3.835	12.107

Kuvassa 26 on piirrettyinä sovitteiden tiheysfunktiot.



Kuva 26: EkspONENTTI-POLYNOMINEN KVANTIILISEKOITUKSEN SOVITTEIDEN TIHEYSFUNKTIOT kolmessa eri ikäryhmässä.

Vastaavat mediaanien erojen estimaatit, keskivirheet ja Bonferroni-korjatut p-arvot ovat taulukossa 14. Tulokset ovat samanlaisia, mutta tämä kvantiilisekoitus

antaa selvästi suuremmat keskivirheet kuin lognormaali-polynominen, joten p-arvot tulevat suuremmiksi. Erityisesti vanhimman ikäryhmän sovitejakauman tyyppi vaikuttaa olevan erilainen kuin lognormaali-polynomisessa.

Taulukko 14: Eksponentti-polynomisen kvantiilisekoituksen ikäryhmien mediaanien vertailu.

Erotus	Arvo	sd	p-arvo
$\hat{M}d_{35-54} - \hat{M}d_{15-34}$	2.325	2.874	> 0.999
$\hat{M}d_{55-79} - \hat{M}d_{15-34}$	11.301	4.066	0.005
$\hat{M}d_{55-79} - \hat{M}d_{35-54}$	8.976	3.993	0.0246

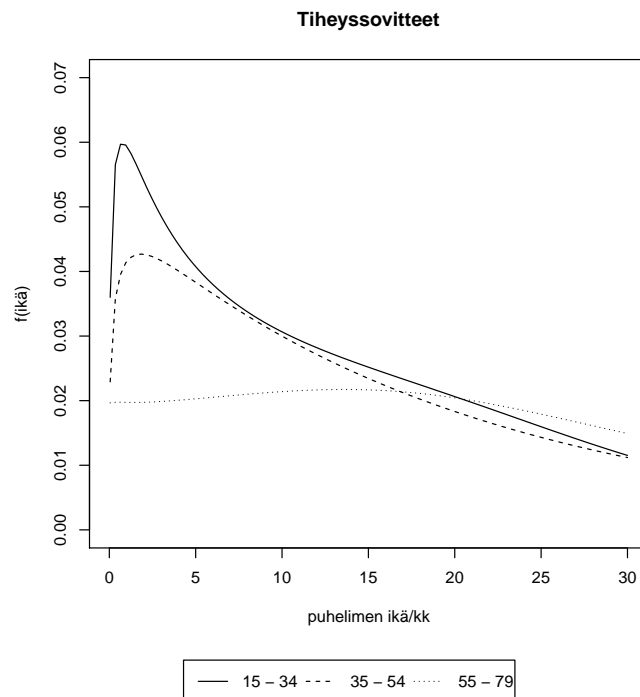
Gamma-polynominen kvantiilisekoitus

Taulukossa 15 on vastaavat gamma-polynomisen kvantiilisekoituksen sovitteet ikäryhmittäin, mediaanin estimaatti ja sen varianssin jackknife-estimaatti.

Taulukko 15: Gamma-polynomisen kvantiilisekoituksen sovitteet ikäryhmittäin.

Estimaatti	15 - 34	35 - 54	55 - 79
\hat{b}	1.513	1.559	1.925
\hat{a}_2	-1.277	-0.189	0.476
\hat{a}_1	1.288	0.077	-1.887
\hat{a}_0	1.419	1.899	3.015
$\hat{M}d$	12.523	14.845	24.226
$\hat{V}ar_{jack}(Md)$	3.383	2.736	14.184

Kuvassa 27 on piirrettynä sovitteiden tiheysfunktiot.



Kuva 27: Gamma-polynomisen kvantiilisekoituksen sovitteiden tiheysfunktiot kolmessa eri ikäryhmässä.

Vastaavat mediaanien erojen estimaatit, keskivirheet ja Bonferroni-korjatut p-arvot ovat taulukossa 16. Tulokset ovat samanlaisia kuin eksponentti-polynomisen kvantiilisekoituksen tapauksessa. Myös jakaumien tyypit eksponentti-polynomisen ja gamma-polynomisen kohdalla ovat hyvin samanlaisia. Huomattavaa kuitenkin on, että jos tarkastellaan eroja 0.05 riskitasolla, niin kahden vanhimman ikäryhmän ero ei ole tällä mallilla merkitsevää.

Taulukko 16: Gamma-polynomisen kvantiilisekoituksen ikäryhmien mediaanien vertailu.

Erotus	Arvo	sd	p-arvo
$\hat{M}d_{35-54} - \hat{M}d_{15-34}$	2.322	2.474	> 0.999
$\hat{M}d_{55-79} - \hat{M}d_{15-34}$	11.703	4.191	0.016
$\hat{M}d_{55-79} - \hat{M}d_{35-54}$	9.381	4.113	0.068

Kaikki kolme mallia löytävät kutakuinkin samanlaiset mediaaniestimaatit, ja vanhimmassa ikäryhmässä käytössä olevien puhelinten ikä vaikuttaa korkeimmalta. Kuitenkin lognormaali-polynomisen kvantiilisekoitus poikkeaa kohtuullisen paljon kahdesta muusta. Edellä huomattiin, että se on useassa tapauksessa heikoimmin sovitettava malli, joten havaitut erot saattavat johtua lognormaali-polynomisen mallin huonosta sopivuudesta.

6.3.2 Alueet

Jaetaan aineisto vastaavasti kahteen osaan henkilön asuinalueen mukaan. Tarkastellaan, että eroaako länsisuomalaisten mediaani muun Suomen mediaanista. Sovitetaan vastaavasti kolme kvantiilisekoitusmallia näihin otoksiin.

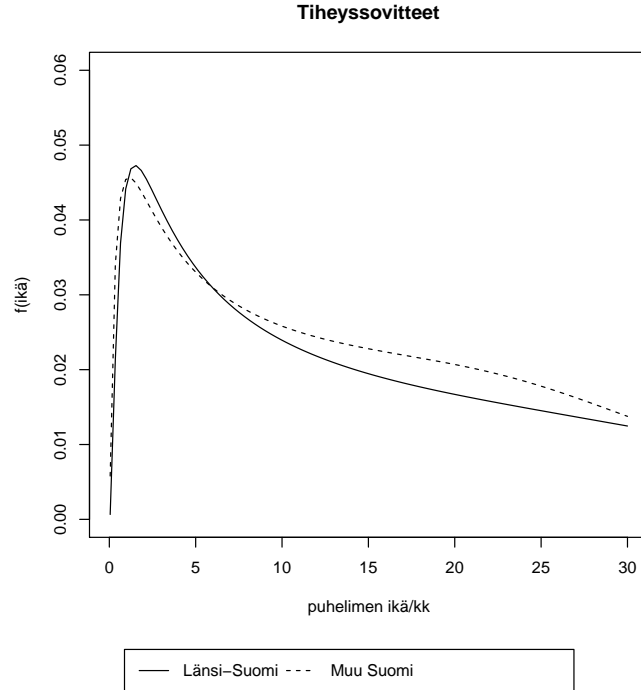
Lognormaali-polynomisen kvantiilisekoitus

Taulukossa 17 on lognormaali-polynomisen kvantiilisekoituksen sovitteet kahdella alueella, mediaanin estimaatti ja sen varianssin jackknife-estimaatti.

Taulukko 17: Lognormaali-polynomisen kvantiilisekoituksen sovitteet alueittain.

Estimaatti	Länsi-Suomi	Muu Suomi
\hat{b}	1.247	1.593
\hat{a}_2	-2.238	-2.478
\hat{a}_1	2.231	1.171
\hat{a}_0	2.350	2.863
$\hat{M}d$	18.284	16.934
$\hat{V}ar_{jack}(Md)$	3.228	1.136

Kuvassa 28 on piirrettynä sovitteiden tiheysfunktiot.



Kuva 28: Lognormaali-polynomisen kvantiilisekoituksen sovitteiden tiheysfunktiot kahdella alueella.

Eri alueiden mediaanien erotukselle saadaan estimaatti, keskivirhe ja p-arvo taulukossa 18. Länsisuomalaisten suurempi mediaani ei tule tilastollisesti merkitseväksi.

Taulukko 18: Lognormaali-polynomisen kvantiilisekoituksen alueiden mediaanien vertailu.

Erotus	Arvo	sd	p-arvo
$\hat{M}d_{\text{Muu Suomi}} - \hat{M}d_{\text{Länsi}}$	-1.350	2.089	0.52

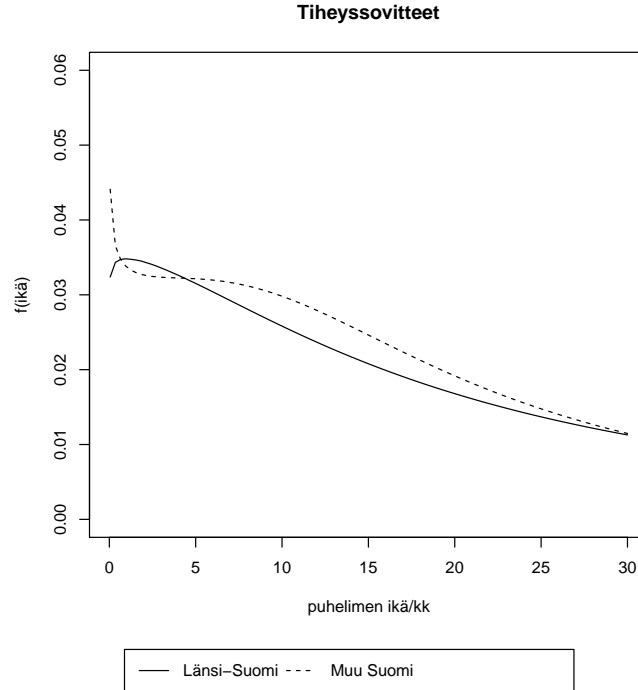
EkspONENTTI-POLYNOMINEN KVANTIILISEKOITUS

Taulukossa 19 on vastaavat eksponentti-polynomisen kvantiilisekoituksen sovitteet alueittain, mediaanin estimaatti ja sen varianssin jackknife-estimaatti. Varianssiestimaateista tulee tällä kvantiilisekoituksella suurempia.

Taulukko 19: EkspONENTTI-POLYNOMINEN KVANTIILISEKOITUKSEN SOVITTEET ALUEITTAIN.

Estimaatti	Länsi-Suomi	Muu Suomi
\hat{b}	0.970	1.119
\hat{a}_2	0.305	0.820
\hat{a}_1	-0.092	-1.409
\hat{a}_0	3.258	3.724
$\hat{M}d$	18.778	16.692
$\hat{V}ar_{jack}(Md)$	6.985	13.589

Kuvassa 29 on piirrettynä sovitteiden tiheysfunktiot. Huomataan, että soviteen tyyppi on varsin erilainen kuin lognormaali-polynomisen kvantiilisekoituksen tapauksessa.



Kuva 29: Eksponentti-polynomisen kvantiilisekoituksen sovitteiden tiheysfunktiot kahdella alueella.

Vastaava mediaanien eron estimaatti, keskivirhe ja p-arvo ovat taulukossa 20. Estimaatti poikkeaa jonkin verran lognormaali-polynomisen antamasta, ja myös sen keskivirhe on suurempi. Ero ei ole tilastollisesti merkitsevä.

Taulukko 20: Eksponentti-polynomisen kvantiilisekoituksen alueiden mediaanien vertailu.

Erotus	Arvo	sd	p-arvo
$\hat{M}d_{\text{Muu Suomi}} - \hat{M}d_{\text{Länsi}}$	-2.089	4.536	0.65

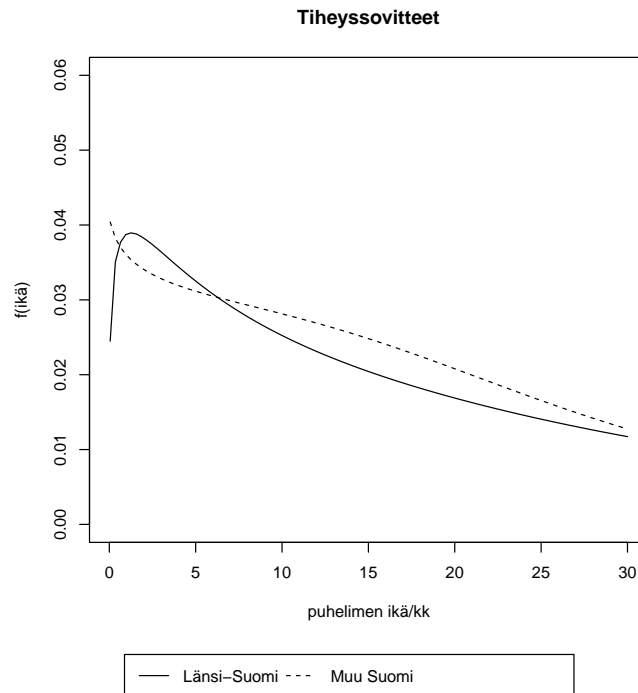
Gamma-polynominen kvantiilisekoitus

Taulukossa 21 on vastaavat gamma-polynomisen kvantiilisekoituksen sovitteet alueittain, mediaanin estimaatti ja sen varianssin jackknife-estimaatti. Huomattavaa on, että Muu Suomi -alueen varianssiestimaatti poikkeaa eksponentti-polynomisen sekoituksen vastaavasta. Tämä saattaa johtua näiden kahden erilaisesta herkkyydestä poikkeaville havainnoille.

Taulukko 21: Gamma-polynomisen kvantiilisekoituksen sovitteet alueittain.

Estimaatti	Länsi-Suomi	Muu Suomi
\hat{b}	1.595	1.967
\hat{a}_2	-0.499	-0.173
\hat{a}_1	0.502	-1.044
\hat{a}_0	1.956	2.383
$\hat{M}d$	18.331	17.050
$\hat{V}ar_{jack}(Md)$	6.608	6.941

Kuvassa 30 on piirrettynä sovitteiden tiheysfunktiot. Nämä ovat kohtuullisen samanlaisia eksponentti-polynomisen mallin sovitteiden kanssa, mutta poikkeavat lognormaali-polynomisen vastaavista.



Kuva 30: Gamma-polynomisen kvantiilisekoituksen sovitteiden tiheysfunktiot kahdella alueella.

Vastaava mediaanien eron estimaatti, keskivirhe ja p-arvo ovat taulukossa 22. Estimaatti on nyt ehkä hieman yllättäen lähempänä lognormaali-polynomisen antamaa. Merkitsevyyttä ei myöskään havaita tämän kvantiilisekoituksen tapauksessa.

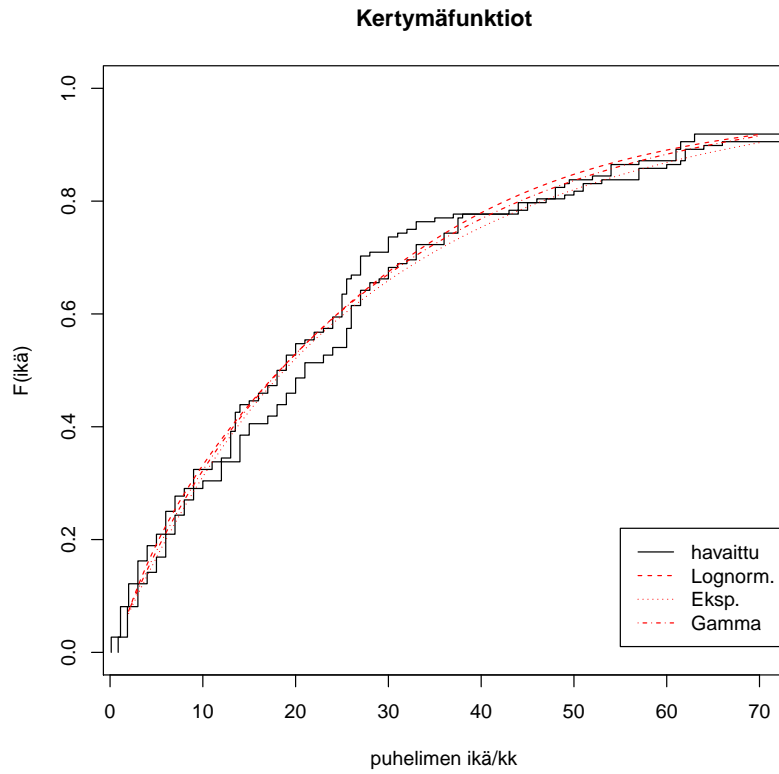
Taulukko 22: Gamma-polynomisen kvantiilisekoituksen alueiden mediaanien vertailu.

Erotus	Arvo	sd	p-arvo
$\hat{M}d_{\text{Muu Suomi}} - \hat{M}d_{\text{Länsi}}$	-1.281	3.681	0.728

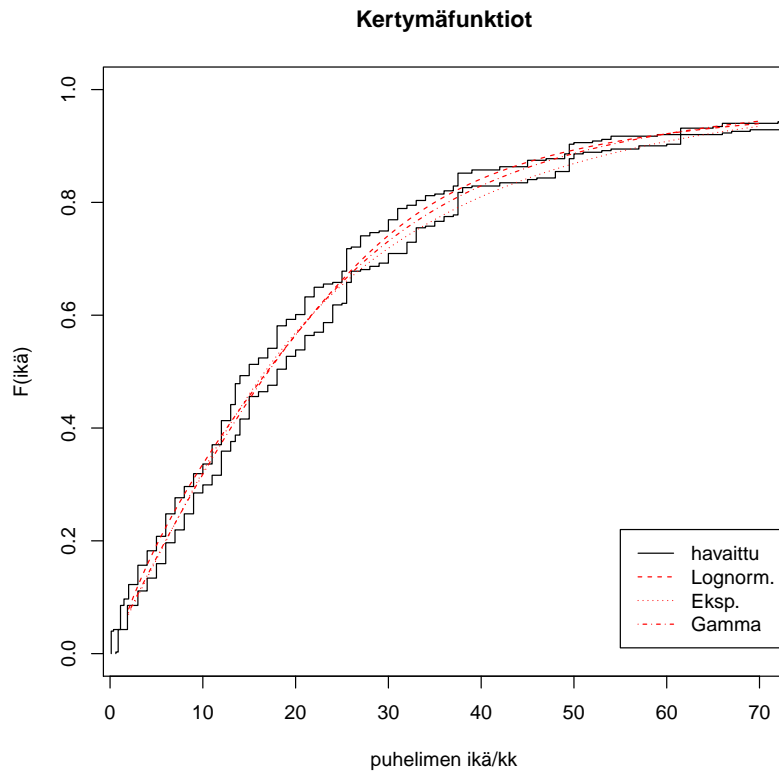
Kaikki kolme mallia antavat saman johtopäätöksen. Alueiden välille ei saada tilastollisesti merkitsevää eroa. Lognormaali-polynomisen kvantiilisekoituksen sovite on jälleen hieman poikkeava kahdesta muusta.

6.4 Diagnostiikkaa

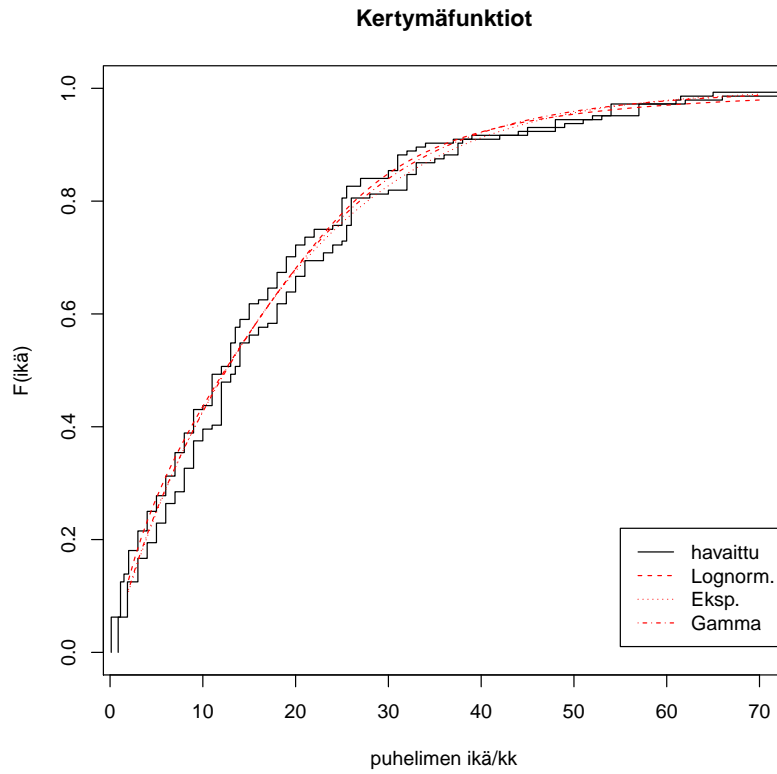
Tarkastellaan sovitettujen mallien sopivuutta kullekin ikäryhmälle ja kummallekin alueelle piirtämällä samaan kuvaan otosvälien ala- ja ylärajojen kertymäfunktiot ja kolmen kvantiilisekoitusmallin kertymäfunktiot. Vertaa artikkeliin (Karvanen et al., 2014). Länsi-Suomen ja Muun Suomen kuvaajat ovat kuvissa 31 ja 32. Eri ikäryhmien kuvaajat ovat kuvissa 33, 34 ja 35.



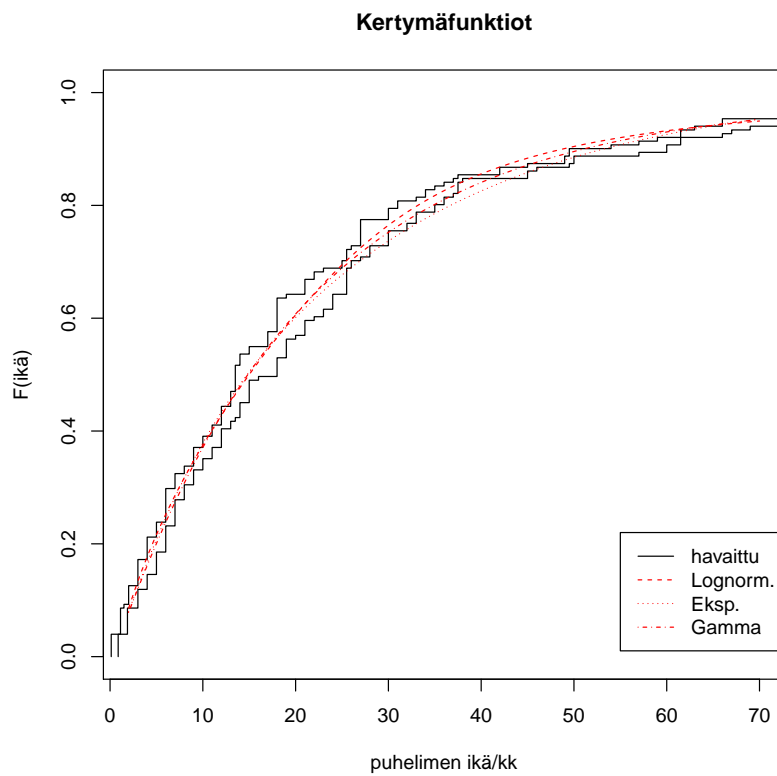
Kuva 31: Länsi-Suomen otoksen kertymäfunktiokuvaajat.



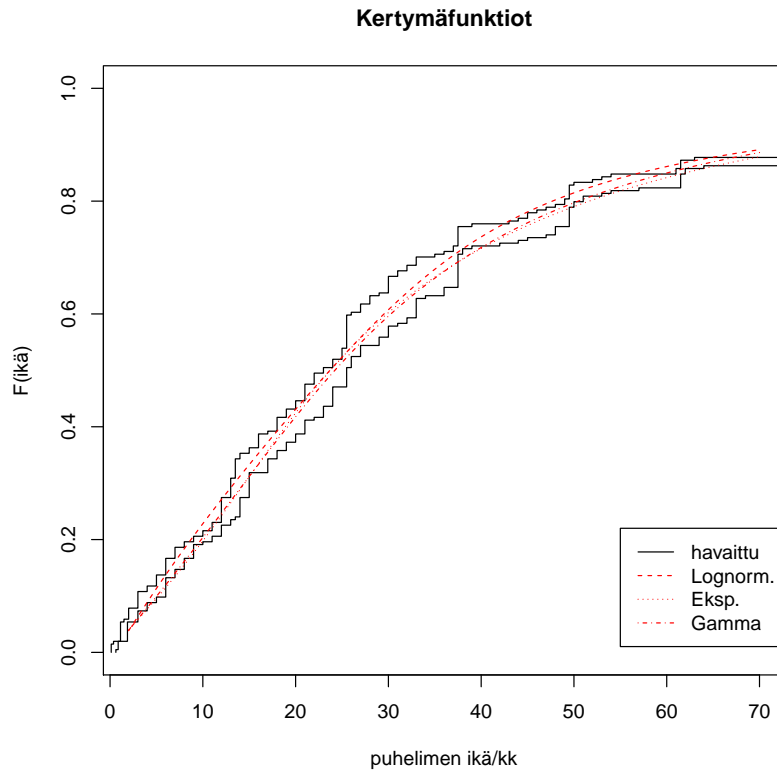
Kuva 32: Muun Suomen otoksen kertymäfunktiokuvaajat.



Kuva 33: Ikäryhmän 15 – 34 otoksen kertymäfunktiokuvaajat.



Kuva 34: Ikäryhmän 35 – 54 otoksen kertymäfunktiokuvaajat.



Kuva 35: Ikäryhmän 55 – 79 otoksen kertymäfunktiokuvaajat.

Mallit näyttävät suurin piirtein sopivilta aineistoon. Länsi-Suomen aineiston kohdalla on mahdollisesti havaittavissa tietynlaista epäsopivuutta, mutta pääasiassa mallin sovitteet kuvaavat hyvin puhelinten ikien jakaumaa.

7 Yhteenveto

Tässä tutkielmassa tarkasteltiin yleistä kvantiilisekoitukseen liittyvää teoriaa ja esiteltiin, kuinka sitä voi soveltaa elinaika-aineistoihin. Havaittiin kolme tapaa, joilla mallin parametrit voi estimoida. Johdettiin kolme erilaista kvantiilisekoitusta, jotka sopivat elinaikadatan mallintamiseen. Verrattiin eri kvantiilisekoitusten sovittavuutta erilaisiin parametrisiin elinaikajakaumiin, erityisesti Weibull-jakaumiin. Lopuksi sovellettiin kvantiilisekoitusmalleja matkapuhelinaineiston analysointiin.

Voidaan todeta, että kvantiilisekoitukset voidaan tarvittaessa muotoilla elinaikojen mallinnukseen sopiviksi. Tässä tutkielmassa on lähestytty ongelmaa konstruomalla muutama yksinkertainen malli ja tarkastelemalla niiden ominaisuuksia lähinnä kuvailevan analyysin keinoin. Syntyneet malliperheet ovat varsin joustavia ja sopivat monessa tilanteessa korvaamaan esimerkiksi Weibull-jakauman. Mallit ovat kuitenkin siinä mielessä heikkoja, että kertoimien tulkinta juurikin elinaikatilanteessa voi olla vaikeaa, eikä mikään kerroin ole suoraan tulkittavissa riskitiheyden tyypiksi, kuten Weibull-jakaumassa. Mallit ovat kuitenkin loppujen lopuksi ihan käyttökelpoisia, varsinkin Weibull-mallin laajentava eksponentti-polynominen kvantiilisekoitus, joka voi olla harkittavissa oleva vaihtoehto, jos tavallinen Weibull-malli ei ole tarpeeksi joustava.

Tällaisten kvantiilisekoitusjakaumien tarkempien ominaisuuksien tutkiminen ja edelleen laajentaminen toisille jakaumille saattaisi kaivata jatkotutkimusta. Mitä tapahtuu, jos neljän kertoimen parametrisointia rajoitetaan? Olisiko mahdollista soveltaa kvantiilisekoituksia elinaikoihin muuten kuin eksponenttisen linkin kautta? Miten malliin voisi liittää selittäviä muuttujia?

Kiitokset tästä työstä erityisesti polynomiset kvantiilisekoitukset kehittäneelle ohjaaja Juha Karvaselle, jolta koko aiheen alun perin sain ja jolta olen saanut ideat suurimpaan osaan tässäkin tutkielmassa käsitellyistä asioista. Lisäksi kiitokset Jukka Nyblomille sekä tilastotieteen pro gradu -seminaarien osanottajille talvena 2013 – 2014 viitteistä, vihjeistä ja kommentteista, joita pystyin ottamaan huomioon vielä ihan loppuvaiheessakin, sekä lisäksi Salme Kärkkäiselle, joka patisti minua tutkielman tekemisen aloittamiseen niin, että pääsin lopulta aloittamaan vain kohtuullisesti myöhässä.

Viitteet

- Elamir, E. A. H. ja Seheult, A. H. (2003). Trimmed L-moments. *Computational Statistics & Data analysis*, 43(3):299–314.
- Elamir, E. A. H. ja Seheult, A. H. (2004). Exact variance structure of sample L-moments. *Journal of Statistical Planning and Inference*, 124(2):337 – 359.
- Gilchrist, W. G. (2000). *Statistical Modelling with Quantile Functions*. CRC Press, Abingdon.
- Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of Royal Statistical society*, 52(1):105–124.
- Jenkins, S. P. (2005). Survival analysis. <https://www.iser.essex.ac.uk/resources/survival-analysis-with-stata>. Luettu: 9.9.2014.
- Jones, M. (1992). Estimating densities, quantiles, quantile densities and density quantiles. *Annals of the Institute of Statistical Mathematics*, 44(4):721–727.
- Karvanen, J. (2006). Estimation of quantile mixtures via L-moments and trimmed L-moments. *Computational Statistics & Data Analysis*, 51(2):947–959.
- Karvanen, J. (2011). *L-moments and quantile mixtures*. R package version 1.1-3. R Foundation for Statistical Computing.
- Karvanen, J. ja Luoma, L. (2014). Matkapuhelinten ostotutkimus, helmikuu 2013. <https://dvn.jyu.fi/dvn/dv/mobilephone>. Luettu: 9.9.2014.
- Karvanen, J., Rantanen, A., ja Luoma, L. (2014). Survey data and bayesian analysis: a cost-efficient way to estimate customer equity. *Quantitative Marketing and Economics*, 12:305–329.
- Malham, S. J. ja Wiese, A. (2014). Efficient almost-exact Lévy area sampling. *Statistics & Probability Letters*, 88:50 – 55.
- McCool, J. I. (2012). *Using the Weibull distribution: reliability, modeling and inference*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ.
- Saarela, O., Kulathina, S., ja Karvanen, J. (2012). Secondary analysis under cohort sampling designs using conditional likelihood. *Journal of Probability and Statistics*, 2012:37.
- Sillitto, G. P. (1969). Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika*, 56(3):641–650.
- Viglione, A. (2014). *nsRFA: Non-supervised Regional Frequency Analysis*. R package version 0.7-12.
- Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18:293–297.

Liite 1: Momentit

Olkoon muuttuja Y sellainen, että $\mathbb{E}(|Y|^p) < \infty$, missä $p \geq 1$. Tällöin $\mathbb{E}(|bY + c|^p) < \infty$.

Perustelu.

$$\begin{aligned} \int |bY + c|^p f(x) dx &\leq \int (|bY| + |c|)^p f(x) dx = \int (|b|^p |Y|^p + \dots + |c|^p) f(x) dx \\ &= \int (|b|^p |Y|^p) f(x) dx + \dots + \int |c|^p f(x) dx < \infty. \end{aligned}$$

Tämä pätee, koska tässä muodostuu polynomi muuttujalle $|Y|$, ja $\int |Y|^q f(x) dx < \infty$, kun $q \leq p$. \square

Olkoon muuttuja Y sellainen, että $\mathbb{E}(|e^{bY}|^p) < \infty$, missä $p \geq 1$. Tällöin $\mathbb{E}(|e^{bY+c}|^p) < \infty$.

Perustelu.

$$\int (e^{bY+c})^p f(x) dx = \int (e^{bY})^p (e^c)^p f(x) dx = (e^c)^p \int (e^{bY})^p f(x) dx < \infty.$$

\square

Muuttujalla e^{bY+c} ei välttämättä ole samoja momenteja äärellisinä kuin muuttujalla e^Y , ellei muuttujan Y momentit generoiva funktio m , jolle $m(t) = \mathbb{E}(e^{tY})$, ole äärellinen kaikilla $t \in \mathbb{R}$. Jos näin kuitenkin on, niin $\mathbb{E}(|e^{bY+c}|^p) = e^{pc} \mathbb{E}(e^{pbY}) = e^{pc} m(pb) < \infty$. Eli tässä tapauksessa myös eksponentoidun kvantiilisekoituksen odotusarvo on aina äärellinen.

Liite 2: Aineisto

ID	Ikä	Alue	Alaraja/kk	Yläraja/kk	ID	Ikä	Alue	Alaraja/kk	Yläraja/kk
1	55-64	Lansi	78	81	98	55-64	Lansi	5	6
2	55-64	Etela	61.5	73.5	99	45-54	Uusimaa	13.5	25.5
3	25-34	Lansi	54	57	100	55-64	Uusimaa	3	6
4	45-54	Etela	105	108	101	55-64	Uusimaa	13	14
5	55-64	PohjoinenIta	12	15	102	35-44	PohjoinenIta	75	78
6	55-64	Etela	21	24	103	25-34	PohjoinenIta	25.5	37.5
7	55-64	Lansi	52	53	104	65-79	Lansi	49	50
8	65-79	Lansi	24	27	105	35-44	PohjoinenIta	3	6
9	65-79	Etela	37.5	49.5	106	55-64	Lansi	44	45
10	55-64	Lansi	11	12	107	45-54	PohjoinenIta	9	10
11	45-54	Lansi	14	15	108	65-79	PohjoinenIta	37.5	49.5
12	35-44	PohjoinenIta	0.125	0.875	109	55-64	Etela	13.5	25.5
13	45-54	Lansi	4	5	110	55-64	PohjoinenIta	10	11
14	65-79	Etela	39	42	111	55-64	Lansi	37	38
15	55-64	PohjoinenIta	12	15	112	55-64	Lansi	18	19
16	55-64	PohjoinenIta	22	23	113	45-54	PohjoinenIta	30	33
17	55-64	Lansi	13	14	114	65-79	PohjoinenIta	49.5	61.5
18	15-24	PohjoinenIta	10	11	115	65-79	Etela	85.5	97.5
19	35-44	Lansi	6	7	116	55-64	PohjoinenIta	6	9
21	25-34	PohjoinenIta	13	14	117	65-79	PohjoinenIta	30	33
22	25-34	Uusimaa	25	26	118	55-64	Etela	32	33
23	55-64	PohjoinenIta	25.5	37.5	119	35-44	Uusimaa	13	14
24	25-34	PohjoinenIta	18	19	120	65-79	Etela	123	126
26	65-79	PohjoinenIta	72	75	121	25-34	Lansi	7	8
27	55-64	Etela	13	14	123	65-79	Etela	12	15
29	45-54	Lansi	2	3	124	45-54	Lansi	22	23
30	25-34	Lansi	30	33	125	65-79	PohjoinenIta	9	12
31	65-79	PohjoinenIta	12	13	126	35-44	Uusimaa	3	4
32	55-64	Etela	1.125	1.875	128	45-54	Etela	30	33
33	45-54	PohjoinenIta	121.5	133.5	129	45-54	Etela	21	24
34	55-64	Lansi	48	51	130	45-54	Etela	1.125	1.875
35	65-79	Lansi	35	36	131	15-24	Lansi	87	90
36	35-44	Lansi	17	18	132	65-79	Lansi	109.5	121.5
37	35-44	PohjoinenIta	7	8	133	65-79	PohjoinenIta	7	8
38	35-44	Etela	31	32	134	35-44	PohjoinenIta	34	35
39	25-34	Uusimaa	24	25	135	15-24	Lansi	1.125	1.875
40	55-64	Lansi	20	21	136	55-64	Uusimaa	37	38
42	25-34	Uusimaa	18	19	137	25-34	Etela	7	8
44	25-34	Etela	25	26	138	45-54	Lansi	9	10
45	65-79	Lansi	145.5	157.5	139	65-79	PohjoinenIta	50	51
46	45-54	Uusimaa	4	5	140	65-79	Lansi	109.5	121.5
47	45-54	PohjoinenIta	66	69	141	45-54	PohjoinenIta	13.5	25.5
48	55-64	Lansi	30	33	142	55-64	PohjoinenIta	7	8
49	55-64	Lansi	102	105	143	35-44	Etela	4	5
50	25-34	PohjoinenIta	19	20	144	65-79	Lansi	43	44
52	55-64	Etela	21	24	145	65-79	PohjoinenIta	37.5	49.5
53	25-34	PohjoinenIta	25	26	146	65-79	Etela	33	36
54	55-64	Lansi	6	7	147	65-79	Lansi	31	32
55	35-44	Lansi	8	9	148	25-34	Lansi	5	6
56	55-64	Lansi	61	62	149	45-54	Uusimaa	27	28
57	55-64	Lansi	33	36	150	45-54	PohjoinenIta	15	16
58	65-79	Lansi	49.5	61.5	151	25-34	Etela	8	9
59	15-24	PohjoinenIta	22	23	152	65-79	Uusimaa	78	79
60	65-79	Lansi	13	14	153	35-44	PohjoinenIta	9	10
61	45-54	Uusimaa	18	19	154	15-24	Etela	1.5	13.5
62	45-54	Etela	6	7	155	65-79	Lansi	3	6
63	35-44	Lansi	1.125	1.875	156	25-34	Etela	25.5	37.5
64	35-44	Etela	13.5	25.5	157	15-24	Uusimaa	14	15
65	55-64	Lansi	25	26	158	55-64	PohjoinenIta	61.5	73.5
66	35-44	Uusimaa	18	21	159	55-64	PohjoinenIta	18	19
67	65-79	Lansi	61	62	160	55-64	Lansi	25	26
68	25-34	PohjoinenIta	25	26	161	45-54	Lansi	27	30
69	55-64	Lansi	13.5	25.5	162	35-44	Lansi	0.125	0.875
70	35-44	Lansi	12	15	163	45-54	Etela	23	24
71	65-79	Lansi	3	4	164	35-44	Lansi	25.5	37.5
72	55-64	PohjoinenIta	0.125	0.875	165	35-44	Uusimaa	12	13
73	25-34	PohjoinenIta	3	4	166	55-64	Etela	18	21
74	55-64	Lansi	30	31	167	25-34	Uusimaa	15	16
75	55-64	PohjoinenIta	54	57	168	45-54	PohjoinenIta	61.5	73.5
76	55-64	Etela	25.5	37.5	169	35-44	Uusimaa	6	7
77	25-34	Uusimaa	17	18	170	65-79	Lansi	2	3
78	25-34	Uusimaa	31	32	171	65-79	Etela	12	15
79	25-34	Uusimaa	52	53	172	45-54	Lansi	61.5	73.5
80	35-44	Etela	12	13	174	15-24	PohjoinenIta	5	6
81	25-34	Lansi	17	18	175	55-64	Lansi	24	27
82	25-34	Etela	4	5	176	55-64	PohjoinenIta	28	29
83	35-44	Uusimaa	17	18	177	45-54	Uusimaa	73.5	85.5
84	55-64	Lansi	6	9	178	35-44	Lansi	13	14
85	45-54	Uusimaa	2	3	179	45-54	PohjoinenIta	3	4
86	45-54	PohjoinenIta	31	32	180	25-34	Lansi	8	9
87	55-64	Uusimaa	0.375	0.625	181	55-64	Lansi	28	29
88	65-79	Uusimaa	21	24	182	55-64	PohjoinenIta	22	23
89	45-54	Uusimaa	2	3	183	25-34	PohjoinenIta	31	32
90	35-44	Lansi	3	4	184	25-34	Lansi	16	17
91	55-64	Uusimaa	13.5	25.5	185	45-54	Lansi	2	3
93	35-44	Etela	27	30	186	65-79	Etela	138	141
94	15-24	Lansi	15	18	187	25-34	Etela	1.125	1.875
95	35-44	PohjoinenIta	18	19	188	45-54	Etela	0.125	0.875
96	35-44	PohjoinenIta	22	23	189	35-44	PohjoinenIta	36	37
97	25-34	Etela	13	14	190	35-44	Uusimaa	4	5

ID	Ikä	Alue	Alaraja/kk	Yläraja/kk	ID	Ikä	Alue	Alaraja/kk	Yläraja/kk
191	45-54	PohjoinenIta	13	14	289	55-64	Uusimaa	20	21
192	55-64	PohjoinenIta	4	5	290	45-54	PohjoinenIta	45	46
193	15-24	Uusimaa	20	21	292	65-79	Uusimaa	21	24
194	45-54	Lansi	27	28	293	45-54	Lansi	27	30
195	65-79	Uusimaa	17	18	294	25-34	Etela	3	6
197	45-54	Uusimaa	37	38	295	25-34	Lansi	6	7
198	35-44	Lansi	18	19	296	25-34	Etela	11	12
199	45-54	Lansi	3	6	297	65-79	Etela	145.5	157.5
200	65-79	Etela	25.5	37.5	298	45-54	Uusimaa	21	24
201	45-54	PohjoinenIta	12	15	299	35-44	Uusimaa	5	6
202	45-54	Lansi	85.5	97.5	300	55-64	Etela	11	12
203	25-34	Etela	0.125	0.875	301	55-64	Lansi	20	21
204	55-64	PohjoinenIta	1.125	1.875	303	55-64	Etela	49.5	61.5
205	55-64	Lansi	2	3	304	45-54	Etela	1.125	1.875
206	35-44	Uusimaa	17	18	305	25-34	Uusimaa	0.125	0.875
207	35-44	PohjoinenIta	6	7	308	25-34	Etela	5	6
208	25-34	Etela	13.5	25.5	309	15-24	Lansi	0.125	0.875
209	55-64	Lansi	113	114	310	25-34	Etela	13.5	25.5
210	45-54	PohjoinenIta	18	19	311	45-54	Uusimaa	3	4
211	45-54	PohjoinenIta	42	45	312	65-79	Etela	1.125	1.875
212	25-34	Uusimaa	6	9	313	45-54	Lansi	3	6
213	35-44	Etela	42	45	314	45-54	Lansi	6	7
214	45-54	Etela	49	50	315	35-44	Lansi	18	19
215	45-54	PohjoinenIta	1.5	13.5	317	35-44	Lansi	54	57
216	65-79	Etela	21	24	318	15-24	Lansi	6	7
217	15-24	PohjoinenIta	8	9	319	35-44	Uusimaa	10	11
218	55-64	Etela	159	162	320	65-79	Lansi	85.5	97.5
220	45-54	Lansi	5	6	321	15-24	Lansi	33	36
221	25-34	Lansi	4	5	323	35-44	Lansi	9	12
222	35-44	Uusimaa	49.5	61.5	324	45-54	PohjoinenIta	49	50
223	35-44	Lansi	63	66	326	45-54	Lansi	25	26
224	15-24	Uusimaa	12	13	327	25-34	PohjoinenIta	2	3
225	25-34	Lansi	1.125	1.875	328	25-34	Lansi	48	49
226	35-44	Lansi	1.125	1.875	329	25-34	Lansi	30	33
227	45-54	Lansi	9	12	330	35-44	PohjoinenIta	21	24
228	45-54	Lansi	7	8	331	65-79	PohjoinenIta	169.5	181.5
229	15-24	Uusimaa	19	20	332	25-34	Lansi	19	20
230	15-24	Lansi	9	12	333	35-44	PohjoinenIta	12	15
231	25-34	PohjoinenIta	2	3	334	25-34	Lansi	48	51
232	35-44	PohjoinenIta	5	6	335	65-79	PohjoinenIta	37.5	49.5
233	35-44	Etela	8	9	336	65-79	PohjoinenIta	61.5	73.5
234	35-44	Uusimaa	33	36	337	35-44	PohjoinenIta	17	18
235	65-79	Uusimaa	25	26	338	35-44	Lansi	7	8
236	15-24	PohjoinenIta	25	26	339	25-34	PohjoinenIta	5	6
237	55-64	Uusimaa	7	8	340	15-24	Uusimaa	7	8
238	25-34	Lansi	27	30	341	25-34	Lansi	44	45
239	25-34	Lansi	61.5	73.5	342	45-54	Lansi	19	20
240	55-64	Etela	19	20	343	45-54	PohjoinenIta	37.5	49.5
241	35-44	Uusimaa	0.125	0.875	344	15-24	Lansi	0.125	0.875
243	25-34	Etela	8	9	345	35-44	PohjoinenIta	14	15
244	35-44	Lansi	1.125	1.875	346	25-34	Lansi	22	23
245	35-44	Lansi	4	5	347	15-24	Lansi	2	3
246	55-64	Lansi	46	47	348	45-54	Lansi	25	26
248	25-34	Uusimaa	34	35	349	25-34	Etela	7	8
249	25-34	Uusimaa	31	32	350	45-54	PohjoinenIta	18	21
250	45-54	Lansi	26	27	351	25-34	Lansi	0.125	0.875
251	45-54	Etela	1.125	1.875	352	25-34	PohjoinenIta	15	18
252	65-79	Uusimaa	25.5	37.5	353	25-34	Uusimaa	39	42
253	45-54	PohjoinenIta	6	7	354	25-34	Uusimaa	11	12
254	15-24	Uusimaa	9	10	355	15-24	PohjoinenIta	21	24
255	25-34	Lansi	13.5	25.5	356	15-24	Lansi	3	4
256	55-64	Lansi	2	3	357	65-79	Lansi	19	20
257	45-54	Etela	13.5	25.5	358	15-24	PohjoinenIta	11	12
258	15-24	Uusimaa	25.5	37.5	360	25-34	PohjoinenIta	6	7
259	15-24	Uusimaa	11	12	361	25-34	Lansi	1.125	1.875
261	55-64	Etela	13	14	363	15-24	PohjoinenIta	4	5
262	15-24	Etela	13.5	25.5	364	25-34	PohjoinenIta	1.5	13.5
263	15-24	Uusimaa	12	13	365	15-24	Lansi	1.125	1.875
264	25-34	PohjoinenIta	3	4	366	15-24	Lansi	11	12
265	15-24	Lansi	27	28	367	15-24	Lansi	61	62
266	15-24	Lansi	7	8	368	15-24	PohjoinenIta	13	14
267	15-24	Lansi	1.125	1.875	370	25-34	Lansi	13	14
269	15-24	Etela	18	19	371	15-24	Lansi	19	20
270	35-44	Etela	0.125	0.875	372	15-24	PohjoinenIta	1.125	1.875
271	55-64	Etela	49	50	373	15-24	PohjoinenIta	9	12
272	15-24	Etela	0.125	0.875	374	15-24	Lansi	20	21
273	55-64	Etela	13	14	375	65-79	Lansi	25.5	37.5
275	35-44	Etela	11	12	376	65-79	Lansi	23	24
276	45-54	Etela	18	21	377	65-79	Lansi	108	111
277	65-79	Lansi	13.5	25.5	378	65-79	Lansi	25.5	37.5
278	15-24	Uusimaa	65	66	379	15-24	Etela	54	57
279	25-34	Uusimaa	11	12	380	15-24	Etela	2	3
280	65-79	Etela	10	11	381	65-79	Etela	30	33
282	65-79	Etela	4	5	382	15-24	Etela	6	9
283	55-64	Etela	6	7	383	65-79	Uusimaa	6	7
284	65-79	Lansi	13.5	25.5	384	25-34	PohjoinenIta	17	18
285	15-24	Lansi	13	14	385	15-24	PohjoinenIta	20	21
286	25-34	Etela	1.125	1.875	386	15-24	Lansi	18	21
287	65-79	Etela	1.125	1.875	387	25-34	Lansi	9	10
288	45-54	Etela	30	33	388	25-34	Uusimaa	15	16

ID	Ikä	Alue	Alaraja/kk	Yläraja/kk	ID	Ikä	Alue	Alaraja/kk	Yläraja/kk
389	65-79	PohjoinenIta	5	6	485	65-79	Uusimaa	109.5	121.5
390	65-79	PohjoinenIta	25.5	37.5	487	65-79	Uusimaa	111	112
391	25-34	Lansi	54	57	488	65-79	Uusimaa	7	8
392	15-24	PohjoinenIta	7	8	489	65-79	Uusimaa	23	24
394	35-44	Lansi	25.5	37.5	490	65-79	Etela	45	48
395	55-64	PohjoinenIta	27	30	491	65-79	PohjoinenIta	30	33
396	65-79	PohjoinenIta	22	23	492	55-64	Uusimaa	25.5	37.5
397	25-34	PohjoinenIta	31	32	493	45-54	Uusimaa	6	9
399	45-54	PohjoinenIta	10	11	494	15-24	PohjoinenIta	2	3
400	55-64	PohjoinenIta	19	20	495	55-64	Uusimaa	25	26
401	55-64	PohjoinenIta	12	13	496	65-79	Uusimaa	37.5	49.5
402	35-44	PohjoinenIta	13.5	25.5	497	35-44	Uusimaa	1.125	1.875
403	65-79	Lansi	78	81	498	55-64	Uusimaa	2	3
404	25-34	PohjoinenIta	2	3	500	65-79	Uusimaa	37.5	49.5
405	25-34	PohjoinenIta	9	10	501	55-64	Uusimaa	27	30
406	35-44	PohjoinenIta	25.5	37.5	502	45-54	Uusimaa	49.5	61.5
407	55-64	PohjoinenIta	12	15	503	55-64	Uusimaa	3	4
408	65-79	Lansi	63	64	504	25-34	Uusimaa	3	4
409	25-34	Lansi	25	26	505	65-79	Uusimaa	31	32
410	55-64	PohjoinenIta	9	10	506	65-79	Uusimaa	0.125	0.875
411	65-79	Lansi	24	27	507	65-79	Uusimaa	47	48
412	25-34	Lansi	32	33	508	65-79	Uusimaa	33	34
413	15-24	PohjoinenIta	9	12	509	45-54	Uusimaa	27	30
414	25-34	Lansi	14	15	510	65-79	Uusimaa	28	29
415	25-34	Lansi	21	24	511	65-79	Uusimaa	14	15
416	55-64	PohjoinenIta	5	6	512	55-64	Uusimaa	8	9
417	25-34	Lansi	25	26	513	45-54	Uusimaa	13.5	25.5
418	15-24	PohjoinenIta	11	12	514	45-54	Uusimaa	35	36
419	55-64	Lansi	16	17	515	55-64	Uusimaa	8	9
420	45-54	Lansi	57	60	516	65-79	Uusimaa	49.5	61.5
421	55-64	Lansi	13.5	25.5	517	55-64	Etela	12	15
422	35-44	PohjoinenIta	34	35	518	35-44	Uusimaa	6	9
423	25-34	Lansi	4	5	520	65-79	Uusimaa	21	22
424	65-79	Lansi	169.5	181.5	521	65-79	Uusimaa	18	21
426	65-79	Lansi	13	14	522	25-34	Etela	11	12
427	15-24	Lansi	13	14	523	55-64	Uusimaa	45	48
428	15-24	PohjoinenIta	45	48	524	55-64	Uusimaa	75	78
429	45-54	Uusimaa	18	21	525	45-54	Uusimaa	27	28
430	35-44	PohjoinenIta	59	60	526	45-54	Uusimaa	7	8
431	65-79	PohjoinenIta	181.5	193.5	527	45-54	Uusimaa	0.125	0.875
432	25-34	PohjoinenIta	0.125	0.875	528	45-54	Uusimaa	14	15
433	35-44	PohjoinenIta	10	11	529	55-64	Uusimaa	13.5	25.5
434	65-79	PohjoinenIta	25.5	37.5	530	65-79	Uusimaa	15	18
435	35-44	PohjoinenIta	6	7	531	65-79	Uusimaa	0.125	0.875
436	65-79	PohjoinenIta	16	17	532	65-79	Uusimaa	91	92
437	35-44	PohjoinenIta	5	6	533	65-79	PohjoinenIta	18	19
439	65-79	PohjoinenIta	22	23	534	65-79	Uusimaa	15	18
440	15-24	PohjoinenIta	0.125	0.875	535	35-44	Uusimaa	97.5	109.5
441	45-54	PohjoinenIta	11	12	536	55-64	Uusimaa	25.5	37.5
442	55-64	PohjoinenIta	25.5	37.5					
443	65-79	PohjoinenIta	6	9					
444	65-79	PohjoinenIta	3	6					
445	55-64	PohjoinenIta	73.5	85.5					
446	25-34	PohjoinenIta	0.125	0.875					
447	65-79	Uusimaa	16	17					
448	25-34	PohjoinenIta	1.125	1.875					
449	45-54	PohjoinenIta	2	3					
452	45-54	PohjoinenIta	66	67					
453	55-64	PohjoinenIta	30	33					
454	15-24	PohjoinenIta	4	5					
455	15-24	Uusimaa	37	38					
456	15-24	Etela	8	9					
457	55-64	Uusimaa	25.5	37.5					
458	65-79	Uusimaa	37.5	49.5					
459	55-64	Uusimaa	36	39					
460	45-54	Uusimaa	15	18					
461	65-79	Uusimaa	29	30					
462	35-44	Etela	73.5	85.5					
463	65-79	Uusimaa	1.125	1.875					
464	55-64	Uusimaa	1.125	1.875					
465	55-64	Uusimaa	3	4					
466	35-44	Uusimaa	13.5	25.5					
467	55-64	Uusimaa	27	30					
468	65-79	Uusimaa	16	17					
469	45-54	Uusimaa	13	14					
470	65-79	Uusimaa	85	86					
471	65-79	Uusimaa	32	33					
472	65-79	Uusimaa	16	17					
473	45-54	Uusimaa	11	12					
474	65-79	Uusimaa	49.5	61.5					
475	55-64	Uusimaa	5	6					
476	45-54	Uusimaa	21	22					
477	65-79	Uusimaa	14	15					
478	55-64	Uusimaa	11	12					
479	55-64	Uusimaa	1.5	13.5					
480	55-64	Uusimaa	12	15					
481	65-79	Uusimaa	53	54					
482	55-64	Uusimaa	26	27					
483	55-64	Uusimaa	1.125	1.875					
484	45-54	Uusimaa	4	5					

Liite 3: R-koodia

R-koodissa ekponentti-polynomisen ja gamma-polynomisen kvantiilisekoitusten kvantiilikertymä- ja tiheysfunktiot, jakauman simulointi ja parametrien estimointi L-momenttimenetelmällä. Lognormaali-polynomisen on käytännössä toteutettu R-kirjastossa *Lmoments*.

```
# Eksponentti-polynomisen kvantiilisekoitus

# Mallin logaritmin kvantiilifunktio
# u: kertymäpiste
# param: mallin parametrivektori(b,a2,a1,a0)
qlogExpPol<-function(u,param) {
  param[1]*log(-log(1-u))+param[2]*u^2+param[3]*u+param[4]
}

# Mallin kvantiilifunktio
# u: kertymäpiste
# param: mallin parametrivektori
qExpPol<-function(u,param) {
  exp(qlogExpPol(u,param))
}

# Mallin kertymäfunktio numeerisesti ratkaistuna
# y: piste
# param: parametrivektori
pExpPol<-function(y,param) {
  arvo<-NULL
  n<-length(y)

  for (i in 1:n) {
    arvo[i]<-uniroot(f=function(u,param,y0){qlogExpPol(u,param)-y0},
    interval=c(0,1),param=param,y0=log(y[i]),tol = .Machine$double.eps^0.4)$root

    if(abs(arvo[i]-0)<0.0001 | abs(arvo[i]-1)<0.0001) {
      arvo[i]<-as.numeric((1-arvo[i])<0.5)
    }
  }

  return(arvo)
}

# Jakauman tiheysfunktio
# n: otoksen koko
# param: parametrivektori
dExpPol<-function(y,param) {
  n<-length(y)
  arvo<-vector(length=n)
  for (i in 1:n) {
    u<-pExpPol(y[i],param)
    arvo[i]<-(-1/y[i])/(-param[1]/((1-u)*log(1-u))+2*param[2]*u+param[3])
  }
  arvo
}

# Satunnaisotos jakaumasta
# n: otoksen koko
# param: parametrivektori
rExpPol<-function(n,param) {
  un<-runif(n)
  qExpPol(un,param)
}

# Estimo eksponentti-polynomisen mallin parametrit L-momenttimenetelmällä
# otoksesta
# dat: otos (X>0)
datEkspPol<-function(dat) {
  require(Lmoments)
  logdat<-log(dat)
  lmom<-as.vector(Lmoments(logdat,rmax=4))
  lEksp<-c(-0.5772157,0.6931472,-0.1177830,0.1042320)
  A<-matrix(c(lEksp[1],1/3,1/2,1,lEksp[2],1/6,1/6,0,
  lEksp[3],1/30,0,0,lEksp[4],0,0,0),byrow=T,nrow=4)
```



```

solve(A)\%*\%lmom
}

# Eksponentti-polynomisen mallin parametrien kovarianssimatriisi
# otoksesta estimoituna
# dat: otos (X>0)
covEkspPol<-function(dat) {
require(Lmoments)
logdat<-log(dat)
lmom<-as.vector(Lmoments(logdat,rmax=4))
lmom.cov<-Lmomcov(logdat,rmax=4,na.rm=FALSE)
lEksp<-c(-0.5772157,0.6931472,-0.1177830,0.1042320)
A<-matrix(c(lEksp[1],1/3,1/2,1,lEksp[2],1/6,1/6,0,
lEksp[3],1/30,0,0,lEksp[4],0,0,0),byrow=T,nrow=4)
A.minus<-solve(A)
A.minus\%*\%lmom.cov\%*\%t(A.minus)
}

# Gamma-polynomisen kvantiilisekoitus

# Mallin logaritmin kvantiilifunktio
# u: kertymäpiste
# param: mallin parametrivektori(b,a2,a1,a0)
qlogGammaPol<-function(u,param) {
param[1]*log(qgamma(u,shape=2,rate=1))+param[2]*u^2+param[3]*u+param[4]
}

# Mallin kvantiilifunktio
# u: kertymäpiste
# param: mallin parametrivektori
qGammaPol<-function(u,param) {
exp(qlogGammaPol(u,param))
}

# Mallin kertymäfunktio numeerisesti ratkaistuna
# y: piste
# param: parametrivektori
pGammaPol<-function(y,param) {
arvo<-NULL
n<-length(y)

for (i in 1:n) {
arvo[i]<-uniroot(f=function(u,param,y0){qlogGammaPol(u,param)-y0},
interval=c(0,1),param=param,y0=log(y[i]),tol = .Machine$double.eps^0.4)$root

if(abs(arvo[i]-0)<0.0001 | abs(arvo[i]-1)<0.0001) {
arvo[i]<-as.numeric((1-arvo[i])<0.5)
}
}

return(arvo)
}

# Jakauman tiheysfunktio
# n: otoksen koko
# param: parametrivektori
dGammaPol<-function(y,param) {
n<-length(y)
arvo<-vector(length=n)
for (i in 1:n) {
u<-pGammaPol(y[i],param)
y0<-log(qgamma(u,shape=2,rate=1))
arvo[i]<-(1/y[i])/(param[1]/(exp(2*y0-exp(y0)))+2*param[2]*u+param[3])
}
arvo
}

# Satunnaisotos jakaumasta

```

```

# n: otoksen koko
# param: parametrivektori
rGammaPol<-function(n,param) {
un<-runif(n)
qGammaPol(un,param)
}

# Estimoii gamma-polynomisen mallin parametrit L-momenttimenetelmällä
# otoksesta
# dat: otos (X>0)
datGammaPol<-function(dat) {
require(Lmoments)
logdat<-log(dat)
lmom<-as.vector(Lmoments(logdat,rmax=4))
lGamma<-c(0.42301043,0.44336228,-0.05274223,0.06133349)
)
A<-matrix(c(lGamma[1],1/3,1/2,1,lGamma[2],1/6,1/6,0,
lGamma[3],1/30,0,0,lGamma[4],0,0,0),byrow=T,nrow=4)
solve(A)\%*\%lmom
}

# Gamma-polynomisen mallin parametrien kovarianssimatriisi
# otoksesta estimoituna
# dat: otos (X>0)
covGammaPol<-function(dat) {
require(Lmoments)
logdat<-log(dat)
lmom<-as.vector(Lmoments(logdat,rmax=4))
lmom.cov<-Lmomcov(logdat,rmax=4,na.rm=FALSE)
lGamma<-c(0.42301043,0.44336228,-0.05274223,0.06133349)
A<-matrix(c(lGamma[1],1/3,1/2,1,lGamma[2],1/6,1/6,0,
lGamma[3],1/30,0,0,lGamma[4],0,0,0),byrow=T,nrow=4)
A.minus<-solve(A)
A.minus\%*\%lmom.cov\%*\%t(A.minus)
}

```