72

Multiplikatiivisen intensiteettimallin estimointi

Ari Ikonen
30.10.1987
SHV-harjoitustyö

SISÄLLYSLUETTELO

1 JOHDANTO

Käsillä olevan työn tarkoituksena on esitellä ns. multiplikatiivisen intensiteettimallin perusteita ja käyttöä henki- ja eläkevakuutustekniikkaan liittyvissä estimointi- ja testausongelmissa. Teoreettisina apuvälineinä käytetään laskuriprosesseja sekä modernia stokastisten prosessien yleistä teoriaa, erityisesti stokastista integraalilaskentaa. Esitys pyrkii painottumaan käytännöllisiä laskelmia mahdollistaviin tuloksiin, joten tarvittavien teoreettisluontoisten aputulosten täsmälliset todistukset on mahdollisuuksien mukaan ohitettu kirjallisuusviittauksin. Keskeisiä tuloksia valottamaan sekä antamaan selkeän mielikuvan tulosten sovellettavuudesta on mukaan liitetty laskuesimerkkejä, jotka pohjautuvat tietokonesimulointeihin.

Pääasiallisina lähdeteoksina on käytetty stokastisten prosessien yleisen teorian osalta teosta Jacod(1979) (Calcul Stochastique et Problèmes de Martingales) ja laskuriprosessien sekä multiplikatiivisen intensiteettimallin teorian osalta teosta Jacobsen(1982) (Statistical Analysis of Counting Processes). Monet käytännölliset näkökohdat ja päättelyt perustuvat Per Kragh Andersenin ja Ørnulf Borganin matemaattisen tilastotieteen pohjoismaisessa konferenssissa kesällä 1984 Bolkesjøssä pitämään esitelmään "Counting Process Models for Life History Data : a Review" (Andersen & Borgan(1984)).

2 PERUSKÄSITTEITÄ

Seuraavassa oletetaan jatkuvasti, että kaikki käytetyt stokastiset prosessit on määritelty ns. tavalliset ehdot täyttävällä tavalla. Tämä merkitsee sitä, että pohjana on täydellinen todennäköisyyskenttä $(\Omega,\mathscr{L},\mathbf{P})$ sekä äärellinen aikaväli, joka yksinkertaisuuden vuoksi valitaan [0,1]:ksi (käytännön laskuissa on helppo aina suorittaa tarpeellinen ajanmuutos, ellei asia muutoin ole selvä). Käytössä oleva historia $(\mathscr{L}_t)_{t\in[0,1]}$ oletetaan oikealta jatkuvaksi ja täydelliseksi. Tässä oikealta jatkuvuus merkitsee sitä, että

$$\mathscr{L}_{t+} = \bigcap_{s>t} \mathscr{L}_s = \mathscr{L}_t$$

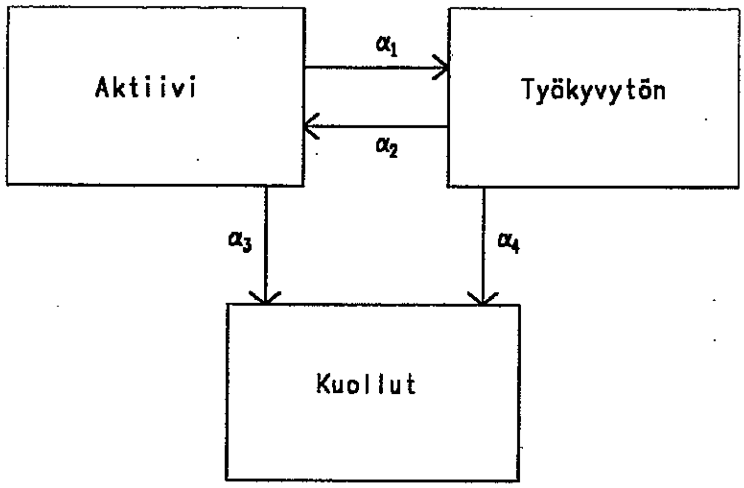
ja täydellisyys sitä, että jokainen $\mathscr{L}_t$ $(t\in[0,1])$ sisältää kaikki $\mathbf{P}$-nollajoukot. Nämä olettamukset ovat luonteeltaan puhtaasti teknisiä eivätkä merkitse käytännöllisiä rajoituksia.

k-ulotteisella laskuriprosessilla

$$N = \{(N_1(t),\dots,N_k(t)) \mid t\in[0,1]\}$$

tarkoitetaan k-ulotteista $(\mathscr{L}_t)$-mukautettua stokastista prosessia, jonka komponenttien N_h $(h=1,\dots,k)$ polut ovat kasvavia oikealta jatkuvia porrasfunktioita, $N_h(0)=0$ $(h=1,\dots,k)$ ja joiden hypyt ovat ykkösen suuruisia. (Kasvavalla takoitetaan tässä kuten jatkossakin funktiota, jonka arvot eivät pienene; jos arvot suurenevat koko ajan, on kyseessä aidosti kasvava funktio.) Lisäksi oletetaan, että todennäköisyys sille, että kaksi komponenttia hyppää samanaikaisesti on nolla, ja jokainen $N_h(t)$ $(h=1,\dots,k$ ja $t\in[0,1])$ on melkein varmasti äärellinen.

Yksinkertainen esimerkki tällaisesta prosessista saadaan henki- ja eläkevakuutuksessa tavallisesta työkyvyttömyysmallista.



Tässä on α_i :llä (i=1,...,4) merkitty tilojen välisiä siirtymäintensiteettejä, joihin palataan myöhemmin. Merkitään $N_i(t)$:llä (i=1,...,4) vastaavien tyyppisten siirtymien aikavälillä [0,t] havaittuja lukumääriä. Nyt

$$N = \{(N_1(t),N_2(t),N_3(t),N_4(t)) \mid t\in[0,1]\}$$

on neliulotteinen laskuriprosessi. Seuraavan luvun tärkeimpänä tavoitteena on käyttökelpoisten estimaattorien etsiminen siirtymäintensiteeteille, kun on havainnoitu vastaavia laskuriprosesseja.

Seuraavassa käytetään jatkuvasti hyväksi martingaalien ja stokastisten integraalien teorian tuloksia ja käsitteistöä. Alla on lyhyt johdatus terminologiaan, tarkemman esityksen täsmällisine todistuksineen voi löytää esimerkiksi teoksen Jacod(1979) luvuista yksi ja kaksi.

Mukautettua stokastista prosessia M, jolle $E(|M_t|)<\infty$ ($t\in[0,1]$) ja jonka polut ovat ns. cadlag-funktioita (so. niillä on kaikissa pisteissä vasemmanpuoleiset raja-arvot ja ne ovat oikealta jatkuvia), kutsutaan martingaaliksi, jos

$$E(M_t|\mathscr{L}_s) = M_s\,,\ 0\leq s\leq t\leq 1,$$

alimartingaaliksi, jos

$$E(M_t|\mathscr{L}_s) \geq M_s\,,\ 0 \leq s \leq t \leq 1,$$

ja ylimartingaaliksi, jos

$$E(M_t|\mathscr{L}_s) \leq M_s\,,\ 0 \leq s \leq t \leq 1,$$

Alimartingaalia M sanotaan neliöintegroituvaksi, jos

$$\sup_{t\in[0,1]} E(M_t^2) < \infty$$

Pysäytyshetki T on satunnaismuuttuja, jolle

$$\{T \leq t\} \in \mathscr{L}_t \text{ kaikilla } t\in[0,1].$$

Jos X on stokastinen prosessi ja T pysäytyshetki, niin pysäytetty prosessi X^T määritellään yhtälöllä

$$X_t^T = X_{T\wedge t}$$

Prosessilla X sanotaan olevan jokin ominaisuus lokaalisti, jos on olemassa kasvava jono $(T_n)_{n\in N}$ pysäytyshetkiä siten, että

$$\mathbf{P}\{T_n \geq t\} \uparrow 1 \text{ kaikilla } t\in[0,1].$$

ja kullakin pysäytetyllä prosessilla X^{T_n} on ko. ominaisuus. Erityisesti seuraavassa käytetään lokaalin martingaalin käsitettä. Huomattakoon heti, että jokainen martingaali on lokaali martingaali.

Prosessia X sanotaan ennustettavaksi, mikäli se on mitallinen ns. ennustettavan σ-algebran $\mathscr{R}$ suhteen. $\mathscr{R}$ puolestaan määritellään muotoa $\{0\} \times A$ $(A\in\mathscr{L}_0)$ ja $]s,1] \times A$ $(A\in\mathscr{L}_s, 0 \leq s \leq 1)$ olevien joukkojen virittämänä σ-algebrana.

(Huomattakoon, että prosessia tarkastellaan kuvauksena $[0,1]\times\Omega\to\mathbf{R}$). Tuleviin tarkoituksiin riittää seuraava yksinkertainen tulos.

Lause 1.1 Jos X on vasemmalta jatkuva (so. polut vasemmalta jatkuvia) ja mukautettu, X on ennustettava.

Todistus: Merkitään

$$X^{(n)} = \mathbf{1}_{\{0\}}X_0 + \sum_{k=1}^{2^n} \mathbf{1}_{](k-1)2^{-n},k2^{-n}]}X_{(k-1)2^{-n}}$$

X on vasemmalta jatkuva, joten

$$X_t(\omega) = \lim_{n\to\infty} X_t^{(n)}(\omega) \text{ kaikilla } (t,\omega)\in[0,1]\times\Omega.$$

Koska yllä olevan summan yhteenlaskettavat ovat $\mathcal{R}$-mitallisia, on X $\mathcal{R}$-mitallinen ja siis ennustettava.□.

Prosessilla X sanotaan olevan kompensaattori Λ , jos X-Λ on lokaali martingaali ja Λ on ennustettava.

Tunnetun Doob-Meyerin esityslauseen mukaan on kaikilla laskuriprosessin komponenteilla N_h yksikäsitteinen kompensaattori Λ_h siten, että prosessit

$$M_h(t) = N_h(t) - \Lambda_h(t) \ , \ h=1,...,k$$

ovat neliöintegroituvia lokaaleja martingaaleja. Tietyin säännöllisyysehdoin (Boel, Varayia, Wong(1975), lauseen 3.2 jälkeiset huomautukset), voidaan prosessit Λ_h valita absoluuttisesti jatkuviksi siten, että on olemassa ennustettavat prosessit λ_h , joille pätee

$$\Lambda_h(t) = \int_0^t \lambda_h(s)ds \ , \ h=1,...,k$$

ja

$$\lambda_h(t+) = \lim_{\Delta t \downarrow 0} \mathbf{P}\{N_h(t+\Delta t)-N_h(t)=1 \mid \mathscr{L}_t\}$$

λ_h:ta kutsutaan N_h:n intensiteettiprosessiksi. Jäljempänä oletetaan koko ajan, että intensiteettiprosessit ovat olemassa kaikille tarkastelluille laskuriprosesseille.

Jos M_1 ja M_2 ovat neliöintegroituvia lokaaleja martingaaleja, niin M_1M_2:lle on olemassa yksikäsitteinen kompensaattoriprosessi, jota merkitään $<M_1,M_2>$:lla. Jos $<M_1,M_2> =0$ kaikilla t, niin M_1 ja M_2 ovat ortogonaalisia.

$<M_1,M_2>$ voidaan tulkita kovariaatioprosessina seuraavasti

$$d<M_1,M_2>(t) = \mathrm{Cov}(dM_1(t),dM_2(t) \mid \mathscr{L}_{t-}).$$

Laskuriprosesseista johdetuille martingaaleille M_h ja M_j (h on erisuuri kuin j) saadaan

$$\begin{aligned} &d<M_h,M_j>(t) \\ &= \mathrm{Cov}(dN_h(t)-\lambda_h(t)dt,dN_j(t)-\lambda_j(t)dt \mid \mathscr{L}_{t-}) \\ &= E(dN_h(t)dN_j(t) \mid \mathscr{L}_{t-}) \\ &= 0. \end{aligned}$$

Tässä on käytetty hyväksi sitä, että λ_h ja λ_j ovat ennustettavia sekä N_h ja N_j eivät hyppää samalla hetkellä.

Jos H_1 ja H_2 ovat ennustettavia ja lokaalisti rajoitettuja prosesseja sekä M_1 ja M_2 ovat neliöintegroituvia lokaaleja martingaaleja, joiden polut ovat lokaalisti rajoitetusti heilahtelevia, niin stokastinen integraali

$$\int_0^t H_h(s)dM_h(s)$$

voidaan määritellä neliöintegroituvana lokaalina martingaalina, jolle

$$<\int_0^t H_1(s)dM_1(s),\int_0^t H_2(s)dM_2(s)> = \int_0^t H_1(s)\,H_2(s)d<M_1,M_2>(s)$$

Tulevaa käyttöä varten todettakoon, että

$$E(\int_0^t H_h(s)dM_h(s)) = 0 \text{ kaikilla t.}$$

3 MULTIPLIKATIIVINEN INTENSITEETTIMALLI JA ESTIMOINTI SIINÄ

Tarkastellaan moniulotteista laskuriprosessia

$$N = \{(N_1(t),\dots,N_k(t)) \mid t \in [0,1]\},$$

jonka intensiteettiprosessi voidaan esittää tulona

$$\lambda_h(t) = \alpha_h(t) Y_h(t)\ ,\ h = 1,\dots,k \text{ ja } t \in [0,1],$$

missä $\alpha_h(t)$:t ovat ei-negatiivisia deterministisiä funktioita ja $Y_h(t)$:t ovat ei-negatiivisia ennustettavia prosesseja. Havainnollinen tulkinta tälle ns. multiplikatiiviselle intensiteettimallille on se, että $\alpha_h(t)$ esittää yksilöllistä siirtymäintensiteettiä ja $Y_h(t)$ kertoo siirtymiselle alttiina olevien yksilöiden määrän juuri ennen hetkeä t.

Seuraavassa pyritään etsimään estimaattoreita suureille $\alpha_h(t)$, kun on seurattu prosesseja $N_h(t)$ ja $Y_h(t)$. Tutkitaan ensiksi suuretta

$$A_h(t) = \int_0^t \alpha_h(s) ds\ .$$

Tämä on nimittäin juuri se suure, joka on suoraan havaintoaineistosta estimoitavissa, kun havaintoaineiston muodostavat havaitut siirtymähetket ja tieto siirtymälle alttiina olevien yksilöiden määrästä. α_h:n itsensä estimaattoreihin palataan tuonnempana. A_h on itsessäänkin varsin hyödyllinen. Esimerkiksi iässä t elossa olevalle maksettavan yksikkösuorituksen pääoma-arvo iässä 0 on korkoutuvuudella δ ja (iästä riippuvalla) kuolevuudella $\mu(t)$ tunnetusti

$$\exp(-(\delta t + \int_0^t \mu(s) ds))\ ,$$

joka riippuu $\mu(t)$:stä vain integroidun muodon välityksellä.

Heuristinen perustelu käytettävälle estimaattorille saadaan kirjoittamalla

$$"dN_h(t) = \alpha_h(t) Y_h(t) dt + \text{kohina}"$$

Luonnollinen estimaattori $A_h(t)$:lle on tämän mukaan

$$\int_0^t (Y_h(s))^{-1} dN_h(s)$$

Kuitenkin huomioon on otettava myös mahdollisuus $Y_h = 0$. Tämän vuoksi määritelläänkin

$$\hat{A}_h(t) = \int_0^t J_h(s)(Y_h(s))^{-1} dN_h(s),$$

missä

$$J_h(t) = \mathbf{1}_{\{Y_h(t) > 0\}} ,$$

ja

$$J_h(t)(Y_h(t))^{-1} = 0, \text{ kun } Y_h(t) = 0.$$

Estimaattoria $\hat{A}$ kutsutaan integroidun intensiteetin Nelson-Aalen estimaattoriksi.

Tämän estimaattorin käyttökelpoisuus perustuu ennenkaikkea sen helppoon laskettavuuteen. Vaikka määritelmässä esiintyvä integraali saattaa ensi näkemältä vaikuttaa hyvinkin hankalalta, palautuu sen laskeminen käytännössä aina äärelliseksi summaksi. Tämän näkee merkitsemällä N_h:n peräkkäisiä hyppäyshetkiä $T_{h1} < T_{h2} < \ldots$. Nyt määritelmä voidaan kirjoittaa vaihtoehtoiseen muotoon

$$\hat{A}_h(t) = \sum_{\{j: T_{hj} \leq t\}} (Y_h(T_{hj}))^{-1} ,$$

mikä on yleisimmin laskuissa käytetty muoto. Tästä nähdään myös, että $\hat{A}_h(t)$ on kasvava ja oikealta jatkuva porrasfunktio, joka hyppää N_h:n havaittuina hyppäyshetkinä T_{hj}. Hypyn suuruus hetkellä T_{hj} on $(Y_h(T_{hj}))^{-1}$.

Voidaan osoittaa, että Nelson-Aalen estimaattori antaa likimääräisesti saman tuloksen kuin klassinen ja intuitiivisesti ilmeinen siirtymäsuhde (occurrence / exposure rate). Jaetaan aikaväli $[0,t]$ osituksella $0=t_0<t_1<\ldots<t_m=t$, joka on niin tiheä, että jokaisella osavälillä havaitaan enintään yksi prosessin N_h hyppy ja $\alpha_h(s)$:ää voidaan pitää likimain vakiona osaväleillä. Merkitään tätä vakioarvoa välillä $]t_j,t_{j+1}]$ α_{hj}:llä. Olkoon Δ osavälin pituus (kaikki samanpituisia). Näillä merkinnöillä saadaan siirtymäsuhteen lausekkeeksi

$$\hat{\alpha}_{hj} = (Y_h(t_j)\Delta)^{-1},$$

jos N_h hyppää ko. välillä ja 0 muuten. Luonnollinen estimaattori suureelle

$$A_h(t) = \sum_j \alpha_{hj}\Delta$$

on

$$\hat{A}_h(t) = \sum_j \hat{\alpha}_{hj}\Delta,$$

mikä on likimain sama kuin tapaukseen liittyvä Nelson-Aalen estimaattori.

Seuraavassa käytetään martingaalien ja stokastisten integraalien teoriaa johdettaessa Nelson-Aalen estimaattorin ominaisuuksia. Määritellään ensiksi apuprosessi

$$A_h^*(t) = \int_0^t \alpha_h(s)J_h(s)ds\ ,$$

Tämä muistuttaa paljon $A_h(t)$:n määritelmää, kun todennäköisyys tapahtumalle

$\{ Y_h(s) = 0$ jollakin $s \leq t\}$ on pieni eli käytettävissä on havaintoja kaikilta ajanhetkiltä. $\hat{A}_h(t)$:n ja $A^*_h(t)$:n erotukselle saadaan lauseke

$$\hat{A}_h(t) - A^*_h(t) = \int_0^t J_h(s)(Y_h(s))^{-1} dM_h(s) ,$$

missä

$$dM_h(s) = dN_h(s) - \alpha_h(s)Y_h(s)ds .$$

$J_h Y_h^{-1}$ on jokaisella h ennustettava prosessi, joten erotus on stokastinen integraali neliöintegroituvan lokaalin martingaalin suhteen ja siis itse neliöintegroituva lokaali martingaali, jonka odotusarvo lisäksi häviää kaikkina ajanhetkinä. Tästä seuraa välittömästi

$$E(\hat{A}_h(t)) = E(A^*_h(t)) , h=1,...,k ,$$

mikä voidaan tulkita kumulatiivisen intensiteetin Nelson-Aalen -estimaattorin likimääräisenä harhattomuutena. Stokastisen integraalin määritelmästä seuraa edelleen, että ko. erotuksella on ennustettava kovariaatioprosessi

$$\begin{aligned}
&< \hat{A}_h - A^*_h, \hat{A}_j - A^*_j > (t) \\
&= \int_0^t J_h(s)(Y_h(s))^{-1} J_j(s)(Y_j(s))^{-1} d < M_h(s), M_j(s) > \\
&= \delta_{hj} \int_0^t J_h(s)(Y_h(s))^{-2} d < M_h(s), M_h(s) > \\
&= \delta_{hj} \int_0^t J_h(s)(Y_h(s))^{-2} \alpha_h(s) Y_h(s) ds \\
&= \delta_{hj} \int_0^t J_h(s)(Y_h(s))^{-1} \alpha_h(s) ds ,
\end{aligned}$$

missä δ_{hj} on Kroneckerin δ-funktio. Lokaalit martingaalit $\hat{A}_h(t)-A^*_h(t)$ $(h=1,...,k)$ ovat siis ortogonaalisia, joten niillä on korreloimattomat lisäykset ja $\hat{A}_h(t)-A^*_h(t)$ on korreloimaton satunnaismuuttujan $\hat{A}_j(s)-A^*_j(s)$ kanssa kaikilla s,t ja $h \neq j$.

Erotusprosessin keskineliöpoikkeama on

$$\tau_h(t) = E(\hat{A}_j(t)-A_j^*(t))^2 = E<\hat{A}_h-A_h^*,\hat{A}_j-A_j^*>(t) .$$

Tälle käytetään estimaattoria

$$\hat{\tau}_h(t) = \int_0^t J_h(s)(Y_h(s))^{-2}dN_h(s),$$

joka voidaan myöskin aina laskea äärellisenä summana. Nyt

$$E(\hat{\tau}_h(t)-\tau_h(t)) = E\int_0^t J_h(s)(Y_h(s))^{-2}dM_h(s) = 0,$$

joten ko. estimaattori on harhaton.

Kumulatiiviselle intensiteetille voidaan johtaa myös luottamusväliestimaattoreita. Käyttökelpoinen tällainen on johdettu esimerkiksi teoksessa Jacobsen(1982) (s. 204), ja sen mukaan 100(1-α)-prosentin luottamusväli $A_h(t)$:lle on

$$A_h(t)\in[\hat{A}_h(t)\pm c_\alpha\, \hat{\tau}_h(1)^{1/2}(1+\hat{\tau}_h(t)(\hat{\tau}_h(1))^{-1})] ,$$

missä c_α on $\sup_{t\in[0,1/2]}|W^0(t)|$:n jakauman ylempi α-fraktiili ja W^0 on tavanomainen Brownin silta joukossa [0,1] (Billingsley(1968)). Taulukoituja arvoja on esimerkiksi teoksessa Hall & Wellner(1980) (s. 141).

Palataan nyt takaisin $\alpha_h(t)$:n estimointiin. Tämä voidaan tehdä tasoittamalla Nelson-Aalen -estimaattorin lisäyksiä. Asetetaan

$$\hat{\alpha}_h(t) = b^{-1}\int_0^t K(b^{-1}(t-s))d\hat{A}_h(s) ,$$

missä K on ns. ydinfunktio. K on rajoitettu ja se häviää välin [-1,1] ulkopuolella sekä sen integraali yli ko. välin on 1. Positiivista parametria b kutsutaan ikkunaksi. Yleinen ja seuraavan luvun laskuissa käytetty ydinfunktio on Epanetshnikovin ydin

$$K(x) = 0.75(1-x^2), |x| \leq 1.$$

Vastaavalla tavalla kuin Nelson-Aalen -estimaattorin tapauksessa voidaan $\hat{\alpha}_h(t)$ laskea äärellisenä summana:

$$\hat{\alpha}_h(t) = b^{-1}\sum_{T_{hj}} K(b^{-1}(t-T_{hj}))(Y_h(T_{hj}))^{-1}$$

4 ESTIMOINTIESIMERKKI JA ARVIOINTIA

Tässä luvussa pyritään esittelemään edellisten tulosten käytännöllistä soveltuvuutta. Tätä varten simuloidaan tietokoneella tilannetta, jossa iältään vaihtelevaa ihmisjoukkoa tarkkaillaan viiden vuoden ajan ja havaitaan siinä sattuvat kuolemantapaukset, jotka puolestaan on arvottu noudattamaan TEL:n kuolevuusperustetta. Havaintotuloksiin sovelletaan edellisen luvun laskentamenettelyjä, ja koska tällä kertaa tiedämme ennakolta "oikean" kuolemisintensiteetin, saamme käsityksen menetelmän tehokkuudesta.

TEL:n kuolevuusperuste on määritelty yksilöllisenä kuolemisintensiteettinä, joka on iästä t riippuvainen:

$$\mu(t) = (a1)\exp((a2)(t+(b2))) ,$$

missä

$$(a1) = 0.00005$$
$$(a2) = 0.095$$
$$(b2) = -3 \text{ miehillä}$$
$$(b2) = -12 \text{ naisilla}$$

Yhdistämällä vakioita toisin voidaan kuolevuusintensiteetti ilmoitaa muodossa:

$$\mu(t) = a\exp(bt) ,$$

missä

$$a = (a1)\exp((a2)(b2))$$
$$b = (a2)$$

lässä T elossa olevan kuoliniän jakaumafunktio saadaan lausekkeesta:

$$F_T(t) = 1-\exp(-\int_T^t \mu(s)ds), \ t \geq T$$

Sijoittamalla tähän yllä oleva TEL-kuolevuusintensiteetin lauseke saadaan

$$F_T(t) = 1-\exp(ab^{-1}(\exp(bT)-\exp(bt)))$$

Kuolinikää kuvaavien satunnaislukujen generoinnissa tarvitaan tämän käänteisfunktio, joka löytyy ratkaisemalla yhtälö:

$$F_T(t) = u$$

t:n suhteen. Tämä johtaa tulokseen:

$$F_T^{-1}(u) = b^{-1}\ln(\exp(bT)-ba^{-1}\ln(1-u)) .$$

Nyt saadaan kuolinikää vastaava satunnaisluku muuntamalla välin [0,1] tasaisesta jakumasta saatua satunnaislukua F_T^{-1} :llä.

Tutkittavan perusjoukon alkuperäinen ikärakenne ei vaikuta tuloksiin muutoin kuin havaintoaineiston mahdollisen pienen määrän aiheuttaman epätarkkuuden välityksellä. Muodollisesti tästä voi vakuuttua laskemalla jakaumalle F_T riskisuhteen (engl. failure rate, hazard rate).

$$r_T(t) = f_T(t)(1-F_T(t))^{-1} = a \exp(bt) ,$$

mikä on riippumaton T:stä.

Tutkimuksen perusjoukoksi otettiin 20 000:n hengen joukko, jonka iät valittiin satunnaisesti siten, että ne noudattavat 31.12.1985 TEL-työsuhteessa olleiden ikäjakaumaa (Eläketurvakeskus(1986)). Kullekin yksilölle arvottiin yllä olevan menettelyn mukaan kuolinhetki, joista otettiin kuitenkin huomioon vain ne, jotka sattuivat viiden vuoden kuluessa ikävälille 15-65 vuotta. Kaikkiaan sattui arvonnassa 267 kuolemantapausta tarkastellulla aikavälillä. Vastaavat kumulatiivisen intensiteetin estimoidut arvot kokonaislukuikien kohdilla ovat liitteessä yksi. Vertailun vuoksi on ko. taulukkoon laskettu myös teoreettisesti oikeat arvot. Paremman mielikuvan saamiseksi on aineisto esitetty myös graafisesti saman liitteen toisella sivulla. Lisäksi on liitteeseen piirretty graafiset tulokset kahdesta erillisestä simuloinnista, joissa on ensimmäiseen verrattuna arvottu samalle perusaineistolle uudet kuoliniät.

Tasoittamalla estimaattorin lisäyksiä voitiin myös laskea estimaatit varsinaisen kuolevuusintensiteetin arvoille. Laskussa käytettiin ikkunana arvoa 5 vuotta, kuitenkin siten että ikäalueen reunoilla käytettiin niin pientä ikkunaa, että se aina sopi molemmilta laidoiltaan ikävälille 15-65 vuotta. Tulokset vertailuarvoineen on esitetty liitteessä kaksi. Graafiset esitykset vastaavat samoja simulointikertoja kuin liitteessä yksi.

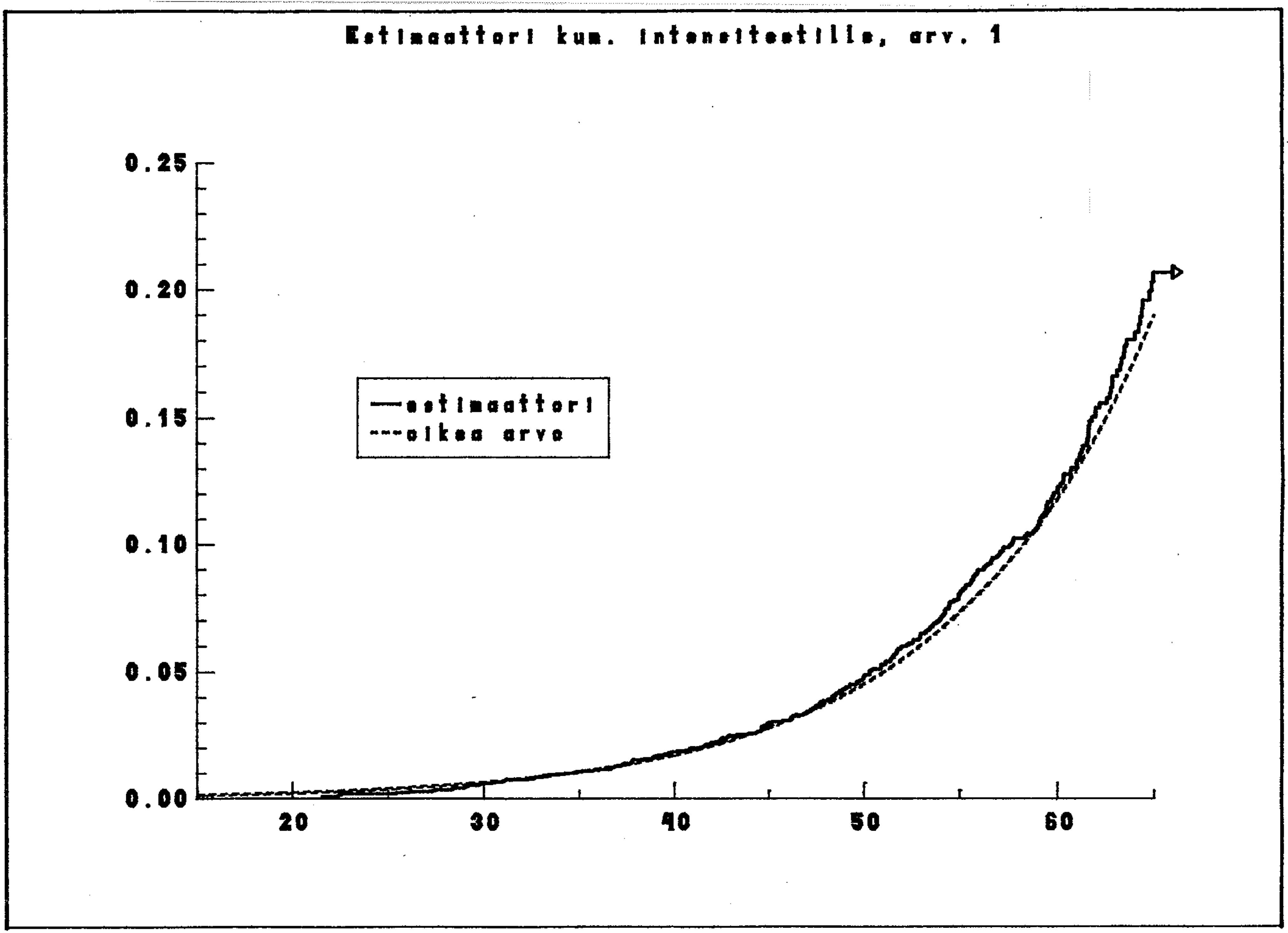
Tuloksia tarkasteltaessa pistää silmään varsinaisen intensiteetin estimaattorien lievä "karkaaminen" ikäskaalan päissä. Tämä aiheutuu tasoituksen hankaluudesta, kun havaintoja ei ole käytettävissä kuin toispuoleisesti. Toisaalta on havaintoaineiston määräkin näissä ikäluokissa pienehkö, mikä luonnollisesti vaikeuttaa tarkkaa estimointia. Vertaamalla eri simulointikerroilta saatuja tuloksia voi todeta, että estimointivirhe on ilmeisestikin satunnainen ja suuruudeltaan melko pieni.

Edellä esitetty menetelmä on yleisesti käyttökelpoinen tapauksissa, joissa intensiteetin multiplikatiivinen hajottaminen ennustettavalla tavalla on mahdollista. Useimmat vakuutustekniset estimointikysymykset ovat helposti tällaisia. Monia estimointimenetelmiä vaikeuttava tutkittavien satunnainen tai systemaattinen katoaminen (esimerkiksi työpaikan vaihtumisen tai muun sellaisen syyn johdosta) ei tuota tälle menetelmälle ongelmia, kunhan vain riskille alttiina olevien lukumäärä kunakin hetkenä tunnetaan ennustettavasti. Vaikka tämä ehto yleensä onkin voimassa, ei sitä voi pitää itsestäänselvyytenä, vaan sen voimassaolo on erikseen mietittävä. Tyypillinen esimerkki saadaan kun sähkölamppujen kestoaikaa testataan kiinteän ajan kestävällä kokeella, jossa käytetään vain yhtä sähkökosketinta. Aina kun lamppu särkyy vaihdetaan uusi tilalle. Koeajan loppuessa jää viimeinen lamppu ehjäksi, joten havainto typistetään (engl. censoring). Harmillista stokastisen integraalilaskennan kannalta on se, että typistyshetki riippuu aikaisemmista ja mahdollisesti pitemmistä kestoajoista, joten riskille alttiina olevien lamppujen lukumäärä ei ole ennustettava prosessi. Käyttämiämme menetelmiä ei voi tässä tilantessa suoraan soveltaa.

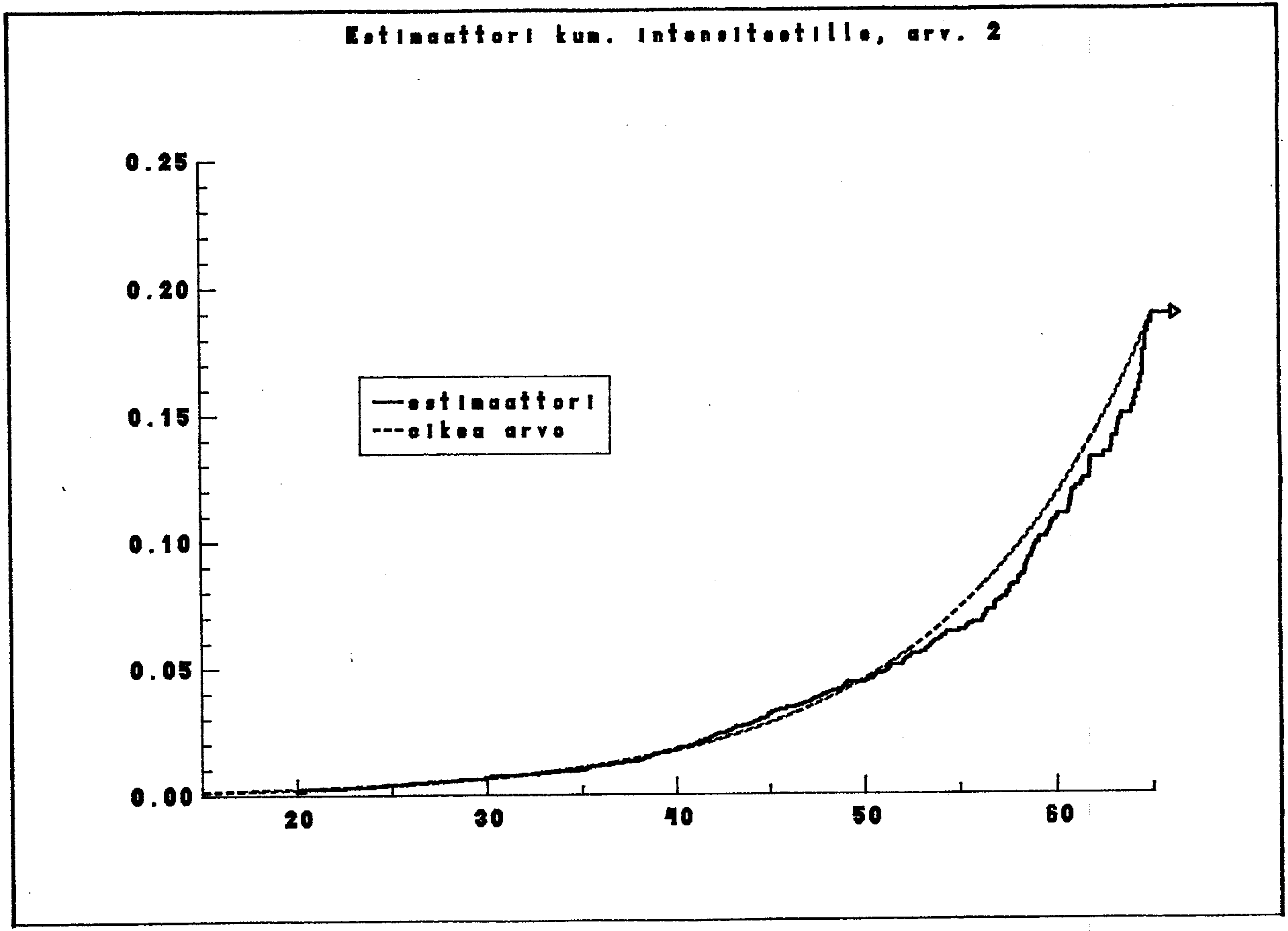
30.10.1987

Estimaattori kumulatiiviselle intensiteetille

Ikä	Estimaatti	Oikea arvo
16	0	0.00141387593
17	0	0.00159422586
18	0	0.001792549257
19	0	0.002010637337
20	0	0.002250459826
21	0	0.002514182749
22	0.00075	0.002804187997
23	0.00135	0.003123094836
24	0.00135	0.003473783565
25	0.00135	0.003859421532
26	0.00233	0.004283491736
27	0.00272	0.004749824292
28	0.00308	0.005262631017
29	0.00379	0.005826543472
30	0.00446	0.006446654798
31	0.00608	0.007128565708
32	0.00701	0.007878435078
33	0.00792	0.008703035572
34	0.00883	0.009609814806
35	0.01005	0.01060696262
36	0.01096	0.01170348505
37	0.01157	0.01290928564
38	0.01307	0.01423525494
39	0.01572	0.01569336882
40	0.01716	0.01729679667
41	0.01886	0.0190600203
42	0.02029	0.02099896477
43	0.02242	0.02313114223
44	0.02537	0.02547581006
45	0.02613	0.02805414479
46	0.03051	0.03088943342
47	0.03140	0.03400728366
48	0.03458	0.03743585529
49	0.03934	0.04120611444
50	0.04372	0.04535211331
51	0.04846	0.04991129767
52	0.05353	0.05492484513
53	0.06065	0.06043803699
54	0.06573	0.06650066723
55	0.07109	0.07316749227
56	0.08093	0.08049872545
57	0.09075	0.08856058094
58	0.09651	0.09742587172
59	0.10354	0.1071746672
60	0.10753	0.1178950165
61	0.12173	0.1296837436
62	0.13193	0.1426473217
63	0.15579	0.1569028351
64	0.16832	0.1725790367
65	0.18626	0.1898175106

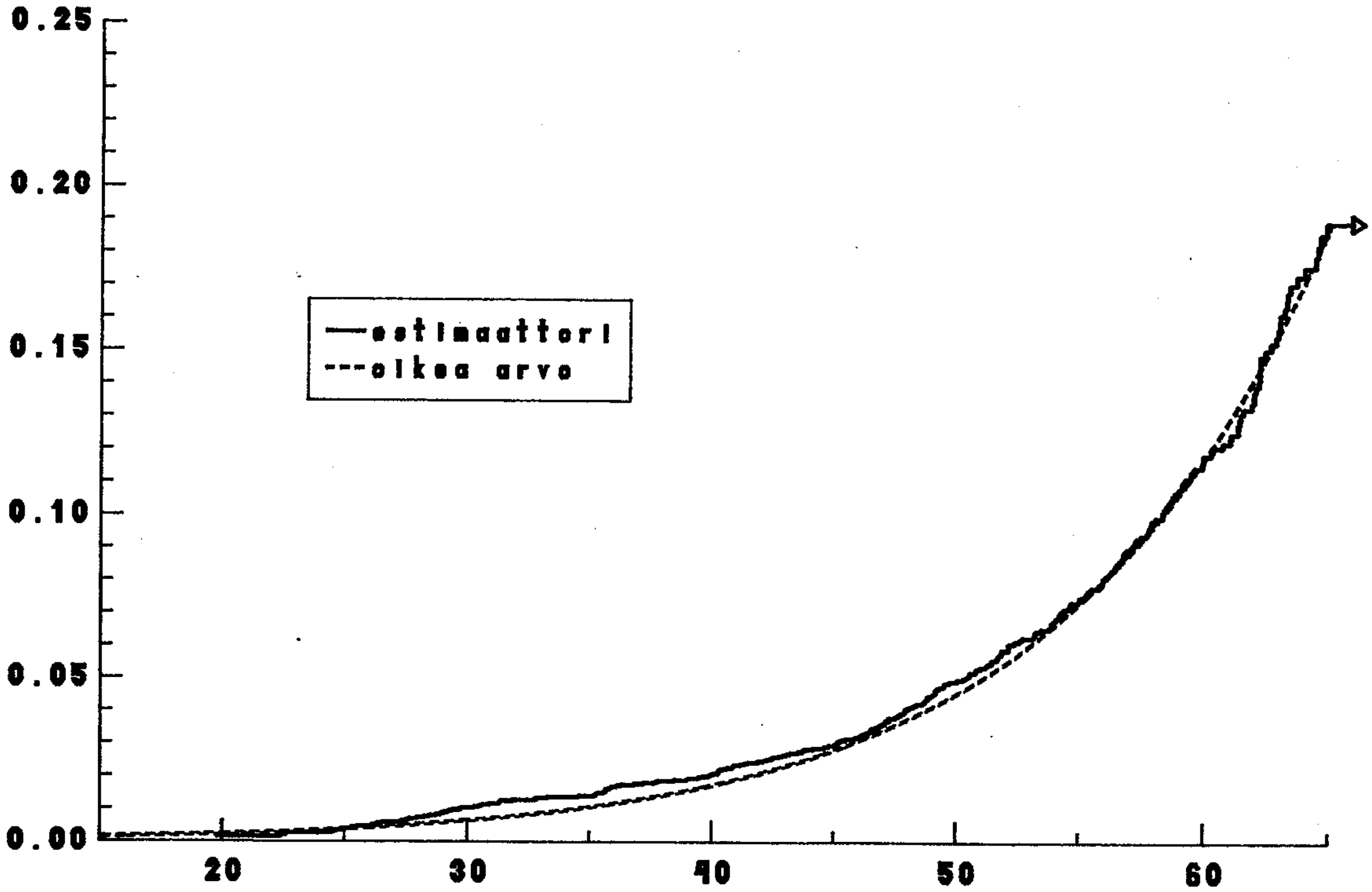


Estimaattori kum. intensiteetille, arv. 1
estimaattori
oikea arvo
0.25
0.20
0.15
0.10
0.05
0.00
20
30
40
50
60



Estimaattori kum. intensiteetille, arv. 2

Estimaattori kum. intensiteetille, arv. 3

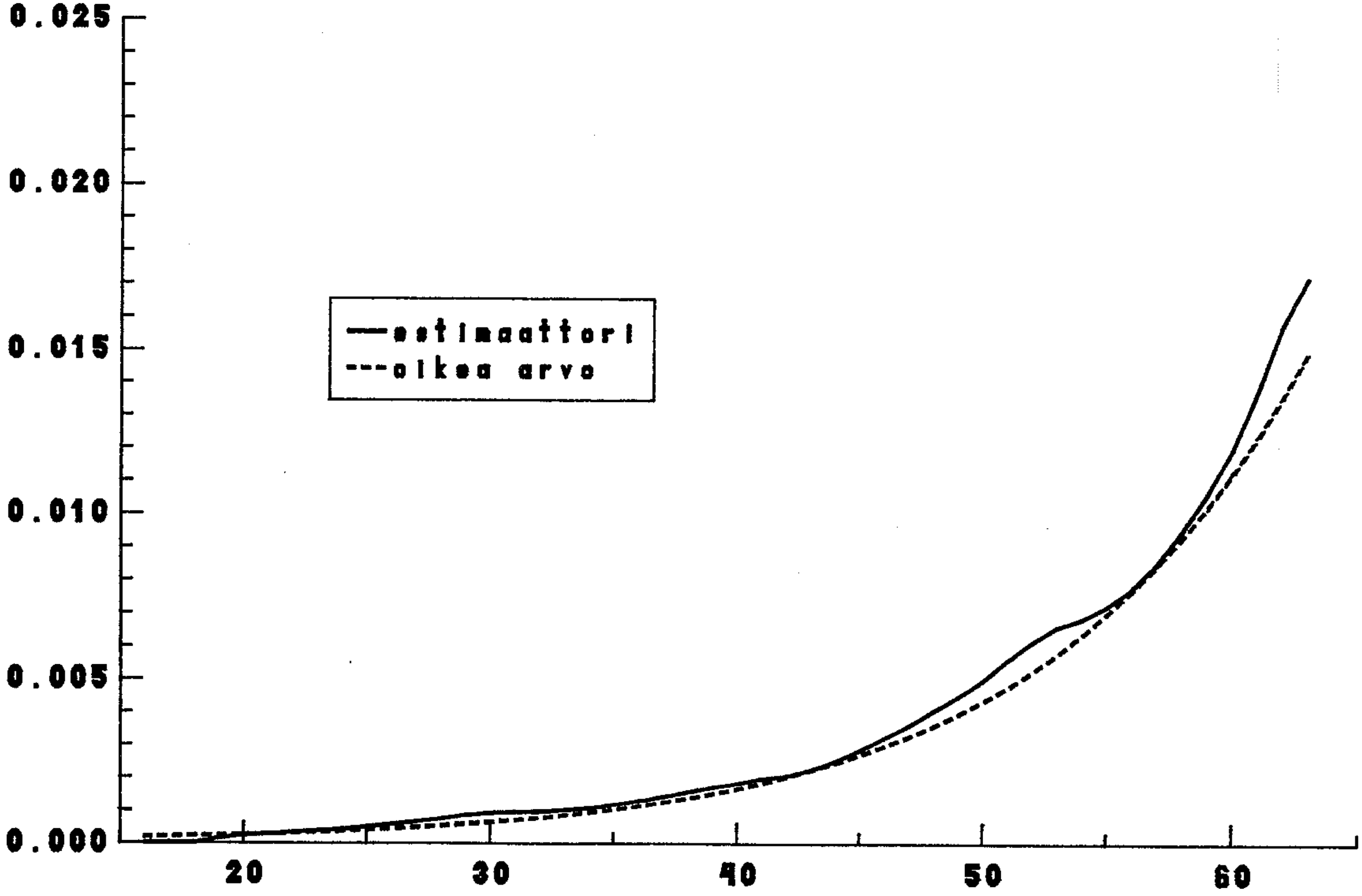


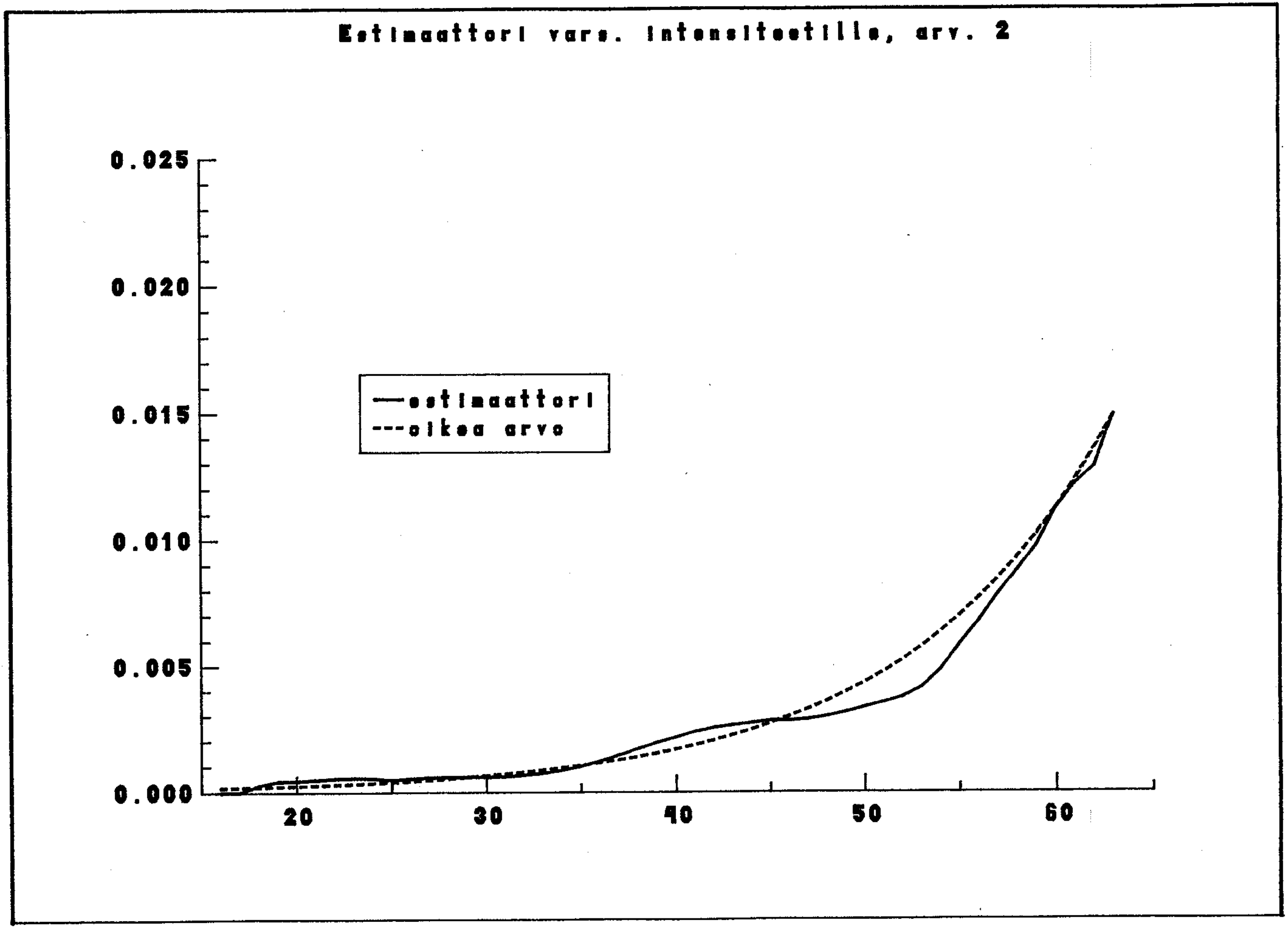
30.10.1987

Estimaattori varsinaiselle intensiteetille

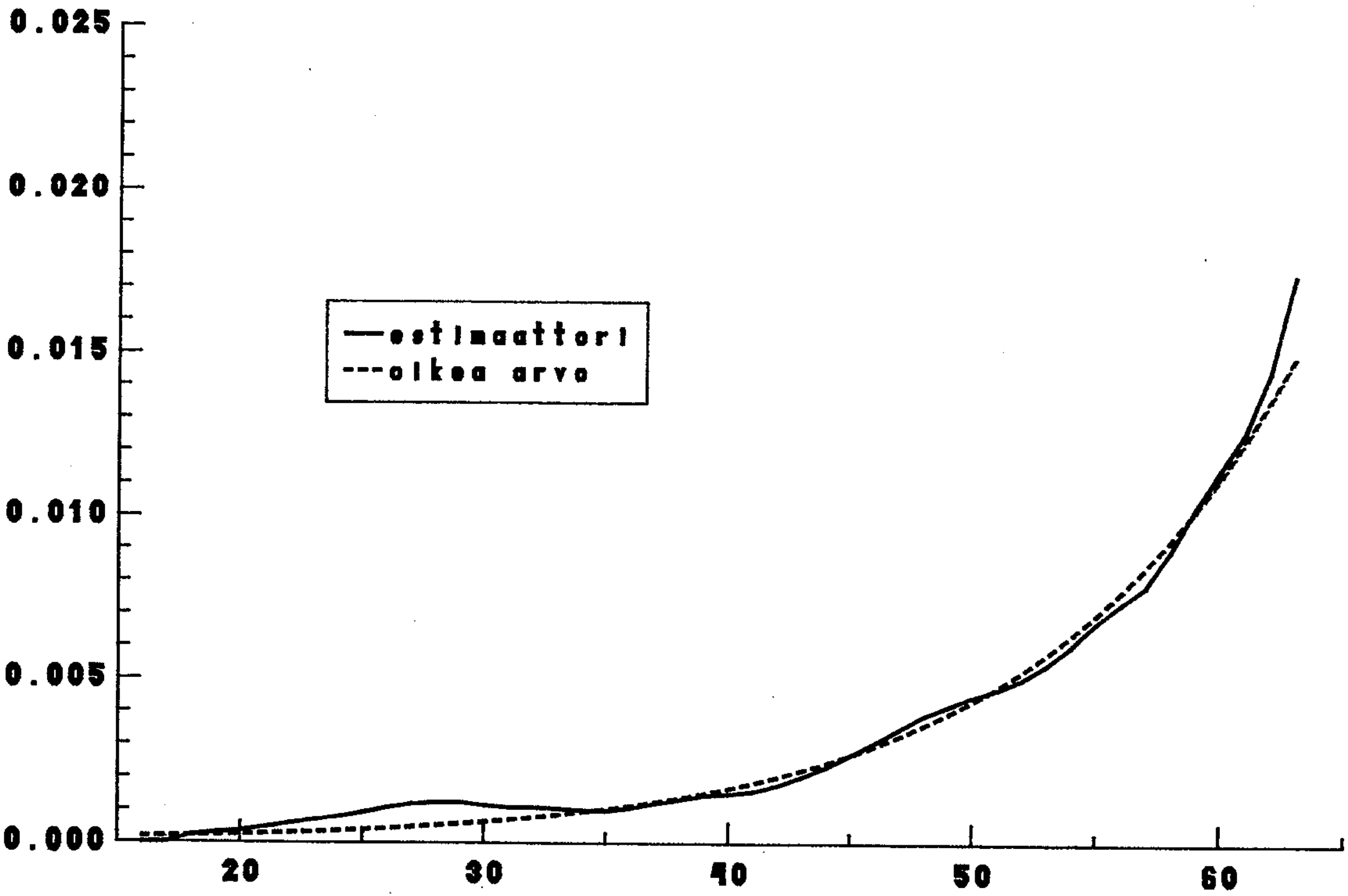
Ikä	Estimaatti	Oikea arvo
16	0	0.000171918926
17	0	0.0001890521694
18	0	0.0002078928921
19	0.00013	0.0002286112598
20	0.00023	0.0002513943962
21	0.00028	0.0002764480739
22	0.00034	0.0003039985724
23	0.00038	0.0003342947221
24	0.00043	0.0003676101514
25	0.00050	0.0004042457582
26	0.00058	0.0004445324277
27	0.00066	0.0004888340205
28	0.00075	0.0005375506593
29	0.00086	0.0005911223426
30	0.00093	0.0006500329185
31	0.00097	0.0007148144549
32	0.00098	0.0007860520451
33	0.00103	0.000864389092
34	0.00109	0.0009505331193
35	0.00119	0.001045262162
36	0.00130	0.001149431792
37	0.00143	0.001263982849
38	0.00158	0.001389949932
39	0.00173	0.001528470751
40	0.00184	0.001680796396
41	0.00199	0.001848302641
42	0.00207	0.002032502366
43	0.00225	0.002235059225
44	0.00252	0.002457802668
45	0.00284	0.002702744468
46	0.00322	0.002972096888
47	0.00361	0.003268292661
48	0.00405	0.003594006965
49	0.00449	0.003952181585
50	0.00498	0.004346051477
51	0.00561	0.004779173992
52	0.00615	0.005255461
53	0.00661	0.005779214226
54	0.00685	0.0063551641
55	0.00721	0.006988512478
56	0.00775	0.007684979631
57	0.00852	0.008450855902
58	0.00949	0.009293058526
59	0.01063	0.0102191941
60	0.01198	0.01123762728
61	0.01376	0.01235755635
62	0.01581	0.01358909627
63	0.01722	0.01494337005
64	0.01852	0.0164326092
65	0	0.01807026422

Estimaattori vars. intensiteetille, arv. 1




Estimaattori vars. intensiteetille, arv. 2
estimaattori
oikea arvo
0.025
0.020
0.015
0.010
0.005
0.000
20
30
40
50
60

Estimaattori vars. intensiteetille, arv. [illegible]



30.10.1987

Kirjallisuusviitteet

- Andersen & Borgan(1984): Counting process models for life history data: a review
  10. matemaattisen tilastotieteen pohjoismainen konferenssi

- Billingsley(1968): Convergence of Probability Measures
  Wiley, New York

- Boel, Varayia, Wong(1975): Martingales on jump processes. I: Representation results. II: Applications
  SIAM Journal of Control 13, 999-1061

- Eläketurvakeskus(1986): Työeläkejärjestelmän tilastollinen vuosikirja 1985, osa II
  Helsinki

- Hall & Wellner(1980): Confidence bands for a survival curve with censored data
  Biometrika 67, 133-143

- Jacobsen(1982): Statistical Analysis of Counting Processes
  Lecture Notes in Statistics 12, Springer Verlag, Berlin

- Jacod(1979): Calcul Stochastique et Problèmes de Martingales
  Lecture Notes in Mathematics 714, Springer Verlag, Berlin