



**TURUN
YLIOPISTO**

TRENDIMUUTOKSEN HAVAITSEMINEN
TYÖKYVYTTÖMYYSELÄKKEELLE SIIRTYVIEN
MÄÄRÄSTÄ SHIRYAEV-ROBERTSIN MENETELMÄLLÄ

Henna Iire

Pro gradu -tutkielma
Kesäkuu 2020

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

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

TURUN YLIOPISTO
Matematiikan ja tilastotieteen laitos

IIRE, HENNA: Trendimuutoksen havaitseminen työkyvyttömyyseläkkeelle siirtyvien määrästä Shiryayev-Robertsin menetelmällä
Pro gradu -tutkielma, 43 s., 11 liites.
Matematiikka
Kesäkuu 2020

Kun tarkasteltavasta aineistosta halutaan esimerkiksi luoda ennusteita tulevaan, on tärkeää huomata mahdolliset tilastollisten ominaisuuksien muutokset ajoissa. Shiryayev-Robertsin menetelmä on yksi muutoksen havaitsemismenetelmistä, ja sitä käsitellään tässä työssä apuna havaitsemaan muutoksia työkyvyttömyyseläkkeelle vuosittain siirtyvien ihmisten lukumäärissä. Etukäteen ei voida tietää kuinka moni joutuu jäämään työkyvyttömyyseläkkeelle, joten lukumäärää täytyy arvioida aiempien vuosien perusteella, ja tässä Shiryayev-Robertsin menetelmästä voisi olla apua.

Työssä käsitellään aluksi itse menetelmää, miten se on määritelty ja minäkalaisia optimaalisuusominaisuuksia sillä on, jonka jälkeen menetelmää käytetään trendimuutoksen havaitsemiseen työkyvyttömyyseläkkeelle siirtyvien lukumäärästä. Menetelmän avulla saatuja havaintoja verrataan siihen mitä tiedetään eläkkeiden myöntämiseen vaikuttavien sääntöjen muutoksista. Erityisesti vuonna 2005 on tapahtunut muutoksia, joiden vaikutuksia voidaan huomata menetelmän avulla. Työssä on keskitytty erityisesti eri ikäryhmiin sekä kahteen suurimpaan sairauspuuryhmään.

Menetelmän huomattiin havaitsevan hyvin muutoksia, joista osa oli selitettävissä sääntöjen muuttumisella. Menetelmä tuotti myös hyviä ennusteita jotka kulkivat samaan suuntaan mihin havainnotkin näyttivät kulkevan. Huomattiin myös havaintojen skaalauksella sekä väärän hälytyksen todennäköisyydellä olevan suuri merkitys menetelmän herkkyyden kannalta.

Asiasanat: Shiryayev-Robertsin menetelmä, trendimuutos, muutoksen havaitseminen, työkyvyttömyyseläke

Sisällys

1 Johdanto	1
2 Shiryaev-Robertsin menetelmä	2
2.1 Optimaalisuusominaisuudet	3
2.2 Kynnysarvo	11
2.3 Poisson-jakauma	12
2.4 Estimoidun intensiteetin muutoskertoimen konsistenssi	14
3 Muutoksen havaitseminen Shiryaev-Robertsin menetelmällä	21
3.1 Shiryaev-Robertsin menetelmä tason muutokselle	21
3.2 Shiryaev-Robertsin menetelmä trendimuutokselle	23
4 Työkyvyttömyyseläkkeelle siirtyneiden tarkastelu	25
4.1 Aineisto	26
4.2 Menetelmä käytännössä	26
4.3 Väärän hälytyksen todennäköisyyden ja skaalauskerroimen vaikutus	29
4.4 Työkyvyttömyyseläkkeellesiirtymisasteen trendimuutos	33
4.5 Ikäryhmät	34
4.6 Sairauspääryhmät	39
5 Loppupäätelmät	43
Liitteet	44
A Ohjelmalistaus	44
Kirjallisuutta	55

1 Johdanto

Joka vuosi ihmisiä sairastuu niin, että he eivät enää kykene työskentelemään täysipäiväisesti tai ollenkaan. Tällaisia tilanteita varten on olemassa työkyvyttömyyseläke, jota näille ihmisille maksetaan. Työkyvyttömyyseläkkeelle siirtyvien lukumäärää ei voida ennalta tarkkaan tietää, mutta sitä voidaan arvioida tarkastelemalla aiempien vuosien lukumääriä, sekä sitä, onko lukumäärä ollut kasvava vai laskeva aiempiin vuosiin nähden. Jos tässä kasvussa tai laskussa tapahtuu ajan myötä muutosta jompaan kumpaan suuntaan, on se hyvä huomioida jotta osataan arvioida paremmin tulevien vuosien lukumääriä. Tässä apuna toimii muutoksen havaitsemismenetelmät, kuten Shiryaev-Robertsin menetelmä, jota tässä tutkielmassa tarkastellaan.

Shiryaev-Robertsin menetelmässä tarkaillaan tiheysfunktioiden avulla havaintoja, ja jos tiheysfunktiossa tapahtuu muutoksia, se pyritään huomaamaan mahdollisimman nopeasti. Luvussa 2 tämä menetelmä esitellään, jonka jälkeen tarkastellaan sen optimaalisuusominaisuuksia sekä kynnysarvon määrittystä. Kynnysarvoa tarvitaan menetelmän pysäyttämiseen. Tarkastellaan lisäksi, miten menetelmä toimii kun ei tiedetä minkälainen tiheysfunktio tulee muutoksen jälkeen olemaan.

Luvussa 3 käydään läpi miten menetelmällä voidaan havaita muutos aineiston tasossa, eli tiheysfunktiossa, tai trendissä, eli siinä miten havainnot kasvavat tai vähenevät ajan kuluessa. Lopuksi luvussa 4 käytetään Shiryaev-Robertsin menetelmää aineistoon, jossa on työkyvyttömyyseläkkeelle siirtyneiden lukumäärät vuosittain. Käydään läpi eri sairauspäyryhmiä sekä ikäluokkia ja katsotaan minkälaisia muutoksia aineistossa havaitaan sekä verrataan havaittuja muutoksia siihen mitä tiedetään työkyvyttömyyseläkkelle siirtymiseen liittyvistä muutoksista.

2 Shiryaev-Robertsin menetelmä

Muutoksen havaitsemisongelmissa tavoitteena on havaita aineiston jakaumassa satunnaisella ajanhetkellä tapahtuva muutos mahdollisimman nopeasti. Satunnaismuuttujat X_n ovat järjestelmän tilasta peräkkäisinä ajanhetkinä tehtäviä havaintoja. Tilan jakaumassa tapahtuu muutos satunnaisella ajanhetkellä T , jota ei voida suoraan havaita. Tehtävänä on päätellä muutoksen tapahtuneen mahdollisimman nopeasti muutoksen jälkeen käyttämällä Shiryaev-Robertsin menetelmää.

Olkoon sarja $\{X_n\}_{n \geq 1}$ peräkkäin tehtyjä tarkkailtavia havaintoja ja olkoon v järjestysnumero ensimmäiselle havainnolle kun muutos on tapahtunut. Tällöin X_v on ensimmäinen havainto muutoksen jälkeen. Olkoon \mathbb{P}_k ja \mathbb{E}_k todennäköisyys ja odotusarvo kun muutos tapahtuu ajanhetkellä $v = k$, jolle $1 \leq k < \infty$. Tällöin \mathbb{P}_∞ ja \mathbb{E}_∞ ovat odotusarvo ja todennäköisyys kun $v = \infty$, eli muutosta ei tapahdu koskaan. Muutoksen havaitsemisprosessi määritellään lopetusajan T avulla. Kaikille $n \geq 0$ joukko $\{T \leq n\}$ kuuluu σ -algebraan $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, joka on n ensimmäisen havainnon luoma σ -algebra.

Jakauman muutos halutaan havaita mahdollisimman nopeasti samalla pitäen väärin hälytysten lukumäärää mahdollisimman pienenä. Yleisimmät tunnusluvut muutoksen havaitsemisessa ovat väärin hälytyksen odotettu havaitsemisnopeus ja havaitsemisviiveen keskiarvo. Ensimmäisellä tarkoitetaan odotettua havaintojen määrää joka aiheuttaa hälytyksen vaikka muutosta ei olisi tapahtunut, toinen luku kertoo odotetun viiveen muutoksen ja sen havaitsemisen välillä. Shiryaev-Robertsin menetelmän tavoitteena on minimoida keskiarvoinen havaitsemisviive rajoittaen samalla väärin hälytyksen odotettua havaitsemisnopeutta.

Tarkastellaan tapausta jossa havainnot ovat ensin riippumattomia ja samoin jakautuneita tiheysfunktiolla f_0 , ja muutoksen jälkeiset havainnot ovat riippumattomia ja samoin jakautuneita tiheysfunktiolla f_1 . Eli oletetaan, että havainnoilla X_n on tiheysfunktio f_0 kun $n \leq v - 1$ ja tiheysfunktio f_1 kun $n \geq v$. Nyt muutoksen hetkestä ehdollinen tiheysfunktio havainnoille (X_1, \dots, X_n) on

$$p(X_1, \dots, X_n | v = k) = \prod_{i=1}^{k-1} f_0(X_i) \cdot \prod_{i=k}^n f_1(X_i), \quad (1)$$

missä $\prod_{i=j}^m f(X_i) = 1$ kun $j > m$.

Shiryaev-Robertsin menetelmä pysähtyy ja tekee hälytyksen lopetushetkellä

$$T(A) = \inf\{n \geq 1 : S_n \geq A\}, \quad \inf\{\emptyset\} = \infty, \quad (2)$$

missä

$$S_n = \sum_{k=1}^n \frac{p(X_1, \dots, X_n | v = k)}{p(X_1, \dots, X_n | v = \infty)} = \sum_{k=1}^n \prod_{i=k}^n \frac{f_1(X_i)}{f_0(X_i)} \quad (3)$$

on Shiryaev-Robertsin sarjan arvo hetkellä n ja $A > 0$ on kynnysarvo joka kontrolloi väriin hälytysten odotettua havaitsemisnopeutta. Shiryaev-Robertsin sarja $S = (S_n)_{n \geq 1}$ on esitetty artikkeleissa [5] ja [6]. Sarjan arvo (3) voidaan myös esittää rekursiivisesti:

$$S_n = (1 + S_{n-1}) \frac{f_1(X_n)}{f_0(X_n)}, \quad n \geq 1, S_0 = 0. \quad (4)$$

Tämän vuoksi klassinen Shiryaev-Robertsin sarja lähtee liikkeelle arvosta 0.[2]

2.1 Optimaalisuusominaisuudet

Shiryaev-Robertsin menetelmällä on useita optimaalisuusominaisuuksia, joista osaa käsitellään tässä kappaleessa.

Pollak tarkasteli huonoimman tapauksen havaitsemisviiveettä, joka on supremum havaitsemisviiveen odotusarvosta ehdollistettuna sillä, että lopetusajankohta on suurempi tai yhtä suuri kuin muutoksen ajankohta,

$$\mathcal{J}_P(T) = \sup_{0 \leq v < \infty} \mathbb{E}_v[T - v | T \geq v]. \quad (5)$$

Huonoin tapaus havaitsemisviiveestä, kun muutos on tapahtunut, tarkoittaa siis sitä että muutos havaitaan todella paljon muutoksen tapahtumisen jälkeen. Pollak yritti löytää optimaalisen menetelmän joka minimoisi arvon (5), jokaisella lopetusajalla T , jolle väärän hälytyksen odotettu havaitsemisajankohta

$$\mathbb{E}_\infty[T] \geq \gamma, \quad (6)$$

missä $\gamma > 0$ on arvo joka päätetään ennen kuin tarkastelu aloitetaan[2]. Väärän hälytyksen odotettu havaitsemisajankohta tarkoittaa siis hetkeä jolloin menetelmä odotusarvoisesti tekee hälytyksen, mutta havaintojen jakaumassa ei ole oikeasti tapahtunut muutosta. Pollakin ideana oli muokata Shiryaev-Robertsin menetelmää satunnaistamalla alkuehto S_0 rekursiivisesta yhtälöstä (4). Tällöin ehdollinen odotusarvo havaitsemisviiveelle $\mathbb{E}_v[T - v | T \geq v]$ olisi riippumaton ajankohdasta v . Pollakin versiossa Shiryaev-Robertsin havaintosarja alkaa siis satunnaisesta pisteestä, joka on otos Shiryaev-Robertsin tunnusluvun S_n quasi-stationaarisesta jakaumasta. Hän todisti, että tällä tavalla satunnaistettu prosessi on asympotoottisesti (kun $\gamma \rightarrow \infty$) optimaalinen

luokassa $o(1)$ kun minimoidaan havaitsemisviiveen odotusarvon supremumia $\mathcal{J}_P(T)[2]$.

Olkoon $B > 0$ ja määritellään

$$Q_B(x) = \lim_{n \rightarrow \infty} \mathbb{P}_\infty[S_n \leq x | T(B) > n]$$

merkitsemään quasi-stationaarista jakaumaa Shiryaev-Robertsin tunnusluvusta. Olkoon $S_n^{Q_B}$ määritelty rekursiivisesti

$$S_n^{Q_B} = (1 + S_{n-1}^{Q_B}) \frac{f_0(x_n)}{f_\infty(X_n)}, \quad n \geq 1, S_0^{Q_B} \sim Q_B, \quad (7)$$

missä $S_0^{Q_B} \sim Q_B$ tarkoittaa, että $S_0^{Q_B}$ on quasi-stationaarisen jakauman Q_B mukaan jakautunut satunnaismuuttuja. Lopetusaika on nyt

$$T_{\text{srp}}(B) = \inf\{n \geq 1 : S_n^{Q_B} \geq B\}, \quad \inf\{\emptyset\} = \infty. \quad (8)$$

Pollak todisti että valitsemalla $B = B_\gamma$ siten, että $\mathbb{E}_\infty[T_{\text{srp}}(B_\gamma)] = \gamma$, saadaan

$$\mathcal{J}_P(T_{\text{srp}}(B_\gamma)) - \inf_{\{T: \mathbb{E}_\infty[T] \geq \gamma\}} \mathcal{J}_P(T) = o(1) \quad \text{kun } \gamma \rightarrow \infty, \quad (9)$$

missä $o(1) \rightarrow 0$ kun $\gamma \rightarrow \infty$. Yhtälöiden (7) ja (8) luomaa prosessia kutsutaan Shiryaev-Roberts-Pollak menetelmäksi.[2]

Koska Shiryaev-Roberts-Pollak menetelmälle pätee $\mathcal{J}_P(T_{\text{srp}}) = \mathbb{E}_0[T_{\text{srp}}] = \mathbb{E}_v[T_{\text{srp}} - v | T_{\text{srp}} \geq v]$ kaikille $v > 0$ voisi epäillä sen olevan tarkalleen optimaalinen jokaisella $\gamma > 0$. Tätä ei kuitenkaan ole pystytty todistamaan. Viimeisimpänä tuloksena Moustakides, Poluchenko ja Tarkakovsky ovat osoittaneet että Shiryaev-Roberts-Pollak menetelmä ei välttämättä ole optimaalinen jokaisella $\gamma > 0$. He ovat selittäneet ongelmaa pohtimalla yleistettyä Shiryaev-Robertsin prosessia, joka lähtee liikkeelle määrätystä pisteestä r . Nyt lopetusaika on määritelty seuraavasti

$$T_{\text{sr}}^r(A) = \inf\{n \geq 1 : S_n^r \geq A\}, \quad \inf\{\emptyset\} = \infty, \quad (10)$$

missä S_n^r saadaan rekursiosta

$$S_n^r = (1 + S_{n-1}^r) \frac{f_0(x_n)}{f_\infty(X_n)}, \quad n \geq 1, \quad S_0^r = r \geq 0. \quad (11)$$

Ratkaisemalla numeerisesti integraaliyhtälöt Moustakides, Poluchenko ja Tartakovsky huomasivat että yleistetty Shiryaev-Robertsin menetelmä on parempi kuin Shiryaev-Roberts-Pollakin menetelmä, eli $\mathbb{E}_v[T_{\text{sr}}^r - v | T_{\text{sr}}^r \geq v] < \mathbb{E}_0[T_{\text{srp}}]$ kaikilla $v > 0$. Näiden tulosten uskotaan olevan vahva todiste sitä

vastaan että Shiryaev-Roberts-Pollakin menetelmä olisi optimaalinen. Tämä todiste ei kuitenkaan ole täysin aukoton, sillä se saattaa sisältää pienen numeerisen virheen[2].

Tarkastellaan seuraavaksi Bayesiläistä ongelmaa jota merkitään $\mathcal{B}(\rho, c)$. Oletetaan arvon v olevan satunnainen ja että sillä on geometrinen jakauma

$$\mathbb{P}[v = k] = \rho(1 - \rho)^{k-1}, \quad k \geq 1,$$

ja tappio lopetusajanhkellä T on 1 jos $T < v$, ja $c(T - v)$ jos $T \geq v$, missä $0 < \rho < 1$ ja $c > 0$ ovat määrättyjä vakioita. Merkitään $\mathbb{P}^\rho[\bullet] = \sum_{k=1}^{\infty} \rho(1 - \rho)^{k-1} \mathbb{P}_k[\bullet]$ ja vastaava odotusarvo on \mathbb{E}^ρ .

Bayesiläinen ongelma $\mathcal{B}(\rho, c)$ on funktion $\varphi_{\rho,c}$ minimointiongelma, missä

$$\varphi_{\rho,c}(T) = \mathbb{P}^\rho[T < v] + c\mathbb{E}^\rho[(T - v)^+]. \quad (12)$$

Funktio (12) kertoo odotetun kustannuksen. Bayesiläinen sääntö tälle ongelmalle saadaan Shiryaevin menetelmästä, joka on lopetusaika

$$T_{\rho,c} = \min\{n \geq 1 : \mathbb{P}^\rho[v \leq n | \mathcal{F}_n] \geq \delta_{\rho,c}\}, \quad (13)$$

missä $0 < \delta_{\rho,c} < 1$ on sopiva kynnyksisarvo [4].

Bayesiläinen minimointiongelma $\mathcal{B}(\rho, c)$ voidaan ratkaista maksimoimalla

$$\frac{1}{\rho}[1 - \varphi_{\rho,c}(T)] = \frac{\mathbb{P}^\rho[T \geq v]}{\rho} - c \frac{\mathbb{E}^\rho[(T - v)^+]}{\rho}, \quad (14)$$

sillä minimointiongelma $\mathcal{B}(\rho, c)$ on ekvivalentti funktion (14) maksimointiongelman kanssa. Lauseen 2.1 todistuksessa näytetään, että mille tahansa lopetusaikalle T

$$\frac{\mathbb{P}^\rho[T \geq v]}{\rho} \xrightarrow{\rho \rightarrow 0} \mathbb{E}_\infty[T],$$

ja

$$\frac{\mathbb{E}^\rho[(T - v)^+]}{\rho} \xrightarrow{\rho \rightarrow 0} \sum_{k=1}^{\infty} \mathbb{E}_k[(T - v)^+].$$

Täten

$$\frac{1}{\rho}[1 - \varphi_{\rho,c}(T)] \xrightarrow{\rho \rightarrow 0} \mathbb{E}_\infty[T] - c \sum_{k=1}^{\infty} \mathbb{E}_k[(T - k)^+],$$

joka on väärän hälytyksen odotetun havaitsemisajankohdan ja kustannuksen, kun hälytys tapahtuu viiveellä, välinen erotus. Muutoksen havaitsemisprosessin tulisi maksimoida tämä erotus. Eli halutaan havaitsemisviiveestä aiheutuvan kustannuksen olevan mahdollisimman pieni ja samaan aikaan väärän

hälytyksen ajankohdan olevan mahdollisimman suuri.

Osoitetaan myös että Shiryaevin prosessi $T_{\rho,c}$ lähestyy Shiryaev-Robertsin prosessia $T(A)$ kun $\rho \rightarrow 0$. Tästä seuraa että integraali $\sum_{k=1}^{\infty} \mathbb{E}_k[(T - k)^+]$ on minimoitu rajoituksen $\mathbb{E}_{\infty}[T] \geq \gamma$ ollessa voimassa.

Lause 2.1. Olkoon A_{γ} valittu siten että $\mathbb{E}_{\infty}[T_{A_{\gamma}}] \geq \gamma$. Tällöin Shiryaev-Robertsin menetelmä, joka on määritelty ehtojen (2) ja (3) mukaan, minimoi termin

$$\sum_{k=1}^{\infty} \mathbb{E}_k[(T - k)^+] \quad (15)$$

jokaisella lopetusajalla T , joka toteuttaa ehdon $\mathbb{E}_{\infty}[T] \geq \gamma$, toisin sanoen

$$\inf_{T \in \Delta_{\gamma}} \sum_{k=1}^{\infty} \mathbb{E}_k[(T - k)^+] = \sum_{k=1}^{\infty} \mathbb{E}_k[(T_{A_{\gamma}} - k)^+] \quad \text{kaikille } \gamma > 0,$$

missä $\Delta_{\gamma} = \{T : \mathbb{E}_{\infty}[T] \geq \gamma\}$.

Todistus. Mietitään Bayesiläistä ongelmaa $\mathcal{B}(\rho, c)$ jossa v on geometrisesti jakautunut parametrilla ρ ja keskimääräinen tappio on kuten yhtälössä (12). Shiryaev on osoittanut että odotettu tappio (12) saavuttaa minimin, kun lopetusaika on yhtälön (13) mukainen. [4]

Merkitään $\prod_{i=1}^0 f_0(X_i) = 1$ ja Bayesin lausetta soveltamalla huomataan, että kun $n \geq 1$

$$\mathbb{P}^{\rho}[v = j | \mathcal{F}_n] = \frac{\prod_{i=1}^{j-1} f_0(X_i) \prod_{i=j}^n f_1(X_i) \rho (1 - \rho)^{j-1}}{\sum_{k=1}^n \prod_{i=1}^{k-1} f_0(X_i) \prod_{i=k}^n f_1(X_i) \rho (1 - \rho)^{k-1} + \prod_{i=1}^n f_0(X_i)},$$

kun $1 \geq j \geq n$. Tämän avulla saadaan, kun $n \geq 1$

$$\begin{aligned}
\mathbb{P}^\rho[v \leq n | \mathcal{F}_n] &= \frac{\mathbb{P}^\rho[v \leq n; \mathcal{F}_n]}{\mathbb{P}^\rho[\mathcal{F}_n]} \\
&= \frac{\sum_{k=1}^n \prod_{i=1}^{k-1} f_0(X_i) \prod_{i=k}^n f_1(X_i) \rho (1-\rho)^{k-1}}{\sum_{k=1}^n \prod_{i=1}^{k-1} f_0(X_i) \prod_{i=k}^n f_1(X_i) \rho (1-\rho)^{k-1} + \prod_{i=1}^n f_0(X_i) (1-\rho)^n} \\
&= \frac{\sum_{k=1}^n \prod_{i=1}^{k-1} f_0(X_i) \prod_{i=k}^n f_1(X_i) (1-\rho)^{k-1}}{\sum_{k=1}^n \prod_{i=1}^{k-1} f_0(X_i) \prod_{i=k}^n f_1(X_i) (1-\rho)^{k-1} + \frac{\prod_{i=1}^n f_0(X_i) (1-\rho)^n}{\rho}} \\
&= \frac{\left(\sum_{k=1}^n \prod_{i=k}^n \frac{f_1(X_i)}{f_0(X_i)} \frac{1}{1-\rho} \right) \prod_{i=1}^n f_0(X_i) (1-\rho)^n}{\left(\sum_{k=1}^n \prod_{i=k}^n \frac{f_1(X_i)}{f_0(X_i)} \frac{1}{1-\rho} \right) \prod_{i=1}^n f_0(X_i) (1-\rho)^n + \frac{\prod_{i=1}^n f_0(X_i) (1-\rho)^n}{\rho}} \\
&= \frac{\sum_{k=1}^n \prod_{i=k}^n \frac{1}{1-\rho} \frac{f_1(X_i)}{f_0(X_i)}}{\sum_{k=1}^n \prod_{i=k}^n \frac{1}{1-\rho} \frac{f_1(X_i)}{f_0(X_i)} + \frac{1}{\rho}} = \frac{S_{\rho,n}}{S_{\rho,n} + \frac{1}{\rho}},
\end{aligned}$$

missä

$$S_{\rho,n} = \sum_{k=1}^n \prod_{i=k}^n \frac{1}{1-\rho} \frac{f_1(X_i)}{f_0(X_i)}.$$

Täten Shiryaevin sääntö voidaan kirjoittaa muodossa

$$T_{\rho,c} = \min\{n \geq 1 : S_{\rho,n} \geq A_{\rho,c}\}, \quad (16)$$

missä $A_{\rho,c} = \frac{1}{\rho}(\delta_{\rho,c}/(1-\delta_{\rho,c}))$. Huomataan nyt että $S_{\rho,n} \xrightarrow{\rho \rightarrow 0} S_n$.

Pollakin mukaan on olemassa vakio $0 < c^* < \infty$ ja sarja $\{\rho_i, c_i\}_{i=1}^\infty$, joille $\rho_i \xrightarrow{i \rightarrow \infty} 0$, $c_i \xrightarrow{i \rightarrow \infty} c^*$, sellaiset että $T_{\rho_i, c_i} \xrightarrow{i \rightarrow \infty} T_{A_{\rho_i, c_i}}$. Lisäksi

$$\limsup_{\rho \rightarrow 0, c \rightarrow c^*} \frac{1 - \varphi_{\rho,c}(T_{\rho,c})}{1 - \varphi_{\rho,c}(T_{A_{\rho,c}})} = 1, \quad (17)$$

missä $\varphi_{\rho,c}(T)$ on odotettu tappio ongelmalle $\mathcal{B}(\rho, c)$ kun lopetusaika on T . [4]

Nyt mille tahansa lopetusajalle T saadaan, että

$$\begin{aligned} \frac{1}{\rho}(1 - \varphi_{\rho,c}(T)) &= \frac{1}{\rho}(1 - \mathbb{P}^\rho[T < v] - c\mathbb{E}^\rho[(T - v)^+]) \\ &= \frac{1}{\rho}(\mathbb{P}^\rho[T \geq v] - c\mathbb{E}^\rho[T - v|T \geq v]\mathbb{P}^\rho[T \geq v]) \\ &= \frac{\mathbb{P}^\rho[T \geq v]}{\rho}(1 - c\mathbb{E}^\rho[T - v|T \geq v]). \end{aligned}$$

Koska

$$\begin{aligned} \frac{\mathbb{P}^\rho[T \geq v]}{\rho} &= \frac{1}{\rho} \sum_{k=1}^{\infty} \mathbb{P}_k[T \geq k] \rho(1 - \rho)^{k-1} \\ &= \sum_{k=1}^{\infty} \mathbb{P}_\infty[T \geq k] (1 - \rho)^{k-1} \\ &\xrightarrow{\rho \rightarrow 0} \sum_{k=1}^{\infty} \mathbb{P}_\infty[T \geq k] = \mathbb{E}_\infty[T] \end{aligned}$$

ja

$$\begin{aligned} \frac{\mathbb{P}^\rho[T \geq v]\mathbb{E}^\rho[T - v|T \geq v]}{\rho} &= \frac{\mathbb{E}^\rho[T - v; T \geq v]}{\rho} \\ &= \frac{1}{\rho} \sum_{k=1}^{\infty} \mathbb{E}_k[T - k; T \geq k] \rho(1 - \rho)^{k-1} \\ &\xrightarrow{\rho \rightarrow 0} \sum_{k=1}^{\infty} \mathbb{E}_k[T - k; T \geq k] \\ &= \sum_{k=1}^{\infty} \mathbb{E}_k[(T - k)^+], \end{aligned}$$

tästä seuraa, että mille tahansa lopetusajalle T , jolle $\mathbb{E}_\infty[T]$ on äärellinen

$$\frac{1}{\rho}(1 - \varphi_{\rho,c}(T)) \xrightarrow{\rho \rightarrow 0} \mathbb{E}_\infty[T] - c \sum_{k=1}^{\infty} \mathbb{E}_k[(T - k)^+].$$

Tämä yhdessä rajoitteen (17) kanssa osoittaa, että Shiryayev-Robertsin menetelmä minimoi termin (15) kaikilla lopetusajoilla, joille pätee $\mathbb{E}_\infty[T] = \gamma$. Huomataan, että jos $\gamma_1 > \gamma$ niin tällöin myös $T_{A_{\gamma_1}}$ on stokastisesti suurempi kuin T_{A_γ} , eli jokainen odotusarvo termissä (15) kasvaa. Tästä johtuen Shiryayev-Robertsin menetelmä minimoi termin (15) luokassa Δ_γ , jolloin lause on todistettu. \square

Lauseesta 2.1 seuraa lause 2.2.

Lause 2.2. Shiryaev-Robertsin menetelmä minimoi termin

$$\frac{\sum_{k=1}^{\infty} \mathbb{E}_k[T - k | T \geq k] \mathbb{P}_{\infty}[T \geq k]}{\sum_{j=1}^{\infty} \mathbb{P}_{\infty}[T \geq j]} \quad (18)$$

jokaisella lopetusajalla T jolle $\mathbb{E}_{\infty}[T] = \gamma$, eli

$$\begin{aligned} & \inf_{\{T: \mathbb{E}_{\infty}[T] = \gamma\}} \frac{\sum_{k=1}^{\infty} \mathbb{E}_k[T - k | T \geq k] \mathbb{P}_{\infty}[T \geq k]}{\sum_{j=1}^{\infty} \mathbb{P}_{\infty}[T \geq j]} \\ &= \frac{\sum_{k=1}^{\infty} \mathbb{E}_k[T_{A_{\gamma}} - k | T_{A_{\gamma}} \geq k] \mathbb{P}_{\infty}[T_{A_{\gamma}} \geq k]}{\sum_{j=1}^{\infty} \mathbb{P}_{\infty}[T_{A_{\gamma}} \geq j]}, \end{aligned}$$

missä kynnyсарvo A_{γ} valitaan siten, että $\mathbb{E}_{\infty}[T_{A_{\gamma}}] = \gamma$.

Todistus. Koska $\sum_{j=1}^{\infty} \mathbb{P}_{\infty}[T \geq j] = \mathbb{E}_{\infty}[T] = \gamma$, nimittäjä termissä (18) on vakio kaikilla lopetusajoilla T . Osoittajalle pätee

$$\begin{aligned} \mathbb{E}_k[T - k | T \geq k] \mathbb{P}_{\infty}[T \geq k] &= \mathbb{E}_k[T - k | T \geq k] \mathbb{P}_k[T \geq k] \\ &= \mathbb{E}_k[T - k; T \geq k] \\ &= \mathbb{E}_k[(T - k)^+]. \end{aligned}$$

Nyt kun sovelletaan lausetta (2.1), on lause (2.2) todistettu. \square

Lauseet (2.1) ja (2.2) ovat hyödyllisiä seuraavien tulosten todistamisessa. Tarkastellaan seuraavaksi Shiryaev-Robertsin menetelmän optimaalisuutta, kun muutos tapahtuu useiden väärin hälytysten jälkeen. Mietitään tilannetta, jossa on erittäin tärkeää havaita muutos niin nopeasti kuin mahdollista, vaikka se tarkoittaisi väärin hälytysten lukumäärän kasvua. Tämä tarkoittaa että muutosajankohta v on suuri verrattuna vakioon γ , joka tässä tapauksessa määrittää odotetun ajanjakson pituuden perättäisten väärin hälytysten välillä.

Olkoot $T_{A_{\gamma}}^{(1)}, T_{A_{\gamma}}^{(2)}, \dots$ riippumattomia peräkkäisiä toistoja satunnaismuuttujasta $T_{A_{\gamma}}$, joka on määritelty yhtälössä (2), eli

$$T_{A_{\gamma}}^{(i)} = \min \left\{ n > \sum_{j=1}^{i-1} T_{A_{\gamma}}^{(j)} : S_n^{(i)} \geq A_{\gamma} \right\} - \sum_{j=1}^{i-1} T_{A_{\gamma}}^{(j)}, \quad (19)$$

missä $T_{A_{\gamma}}^{(0)} = 0$ ja

$$S_n^i = \sum_{k=T_{A_{\gamma}}^{(i-1)}}^n \prod_{j=k}^n \frac{f_1(X_j)}{f_0(X_j)}, \quad \sum_{j=1}^{i-1} T_{A_{\gamma}}^{(j)} < n \leq \sum_{j=1}^i T_{A_{\gamma}}^{(j)}. \quad (20)$$

Näin ollen $S_n^i, n > \sum_{j=1}^{i-1} T_{A_\gamma}^j$ on vain Shiryaev-Robertsin menetelmän tunnusluku, joka on korjattu alkamaan hetkeä järjestykseltään $i - 1$ olevan väärän hälytyksen jälkeen ja jota sovelletaan havaintoihin

$$X_{\sum_{j=1}^{i-1} T_{A_\gamma}^j + 1}, X_{\sum_{j=1}^{i-1} T_{A_\gamma}^j + 2}, \dots$$

Huomioidaan että $\mathbb{E}_\infty[T_{A_\gamma}^i] = \gamma$ kaikilla $i \geq 1$.

Merkitään

$$Q_j = T_{A_\gamma}^{(1)} + T_{A_\gamma}^{(2)} + \dots + T_{A_\gamma}^{(j)}, \quad (21)$$

jokaiselle $j \geq 1$. Tällöin Q_j on j . hälytyksen ajanhetki. Merkitään $J_v = \min\{j \geq 1 : Q_j \geq v\}$, jolloin Q_{J_v} on ajanhetki, kun ajanhetkellä v tapahtunut muutos oikeasti havaitaan, kun on ensin tehty $J_v - 1$ väärää hälytystä.

Seuraavaksi esitetään lause jonka mukaan termin Q_{J_v} määrittämä Shiryaev-Robertsin menetelmä on asympotoottisesti optimaalinen (kun $v \rightarrow \infty$) kaikista muutoksenhavaitsemismenetelmistä, kun tarkastellaan odotettua havaitsemisviivettä $\mathbb{E}_v[Q_{J_v} - v]$, väärin hälytysten välisten ajanjaksojen pituuden odotusarvon ollessa vähintään γ .

Lause 2.3. Olkoon v ajankohta jolloin muutos tapahtuu, ja olkoon $T_{A_\gamma}^{(1)}, T_{A_\gamma}^{(2)}, \dots$ sarja riippumattomia toistoja satunnaismuuttujasta T_{A_γ} jotka on määritelty yhtälön (19) mukaisesti. Olkoon arvot Q_1, Q_2, \dots kuten yhtälössä (21). Olkoon $J_v = \min\{j : Q_j \geq v\}$.

(i) $\lim_{v \rightarrow \infty} \mathbb{E}_v[Q_{J_v}]$ on olemassa.

(ii) Oletetaan että havaitsemisprosessi T , jolle $\mathbb{E}_\infty[T] = \gamma$, tapahtuu toistuvasti. Olkoon T_1, T_2, \dots peräkkäin tapahtuvia havaitsemisprosesseja ja olkoon $W_j = \sum_{i=1}^j T_i$ ja $K_v = \min\{j : W_j \geq v\}$. Tällöin jokaiselle $\gamma > 0$ pätee

$$\lim_{v \rightarrow \infty} \mathbb{E}_v[Q_{J_v}] \leq \lim_{v \rightarrow \infty} \mathbb{E}_v[W_{K_v} - v]. \quad (22)$$

(iii) Kohdan (ii) epäyhtälö (22) pätee kaikille $T \in \Delta_\gamma$, missä $\Delta_\gamma = \{T : \mathbb{E}_\infty[T] \geq \gamma\}$.

Todistus. Todistetaan ensin kohta (i). Artikkelin [4] mukaan arvon $v - Q_{J_v-1}$ jakaumalla on raja-arvo

$$\lim_{v \rightarrow \infty} \mathbb{P}_v[v - Q_{J_v-1} = k] = \frac{\mathbb{P}_\infty[T_{A_\gamma} \geq k]}{\sum_{j=1}^{\infty} \mathbb{P}_\infty[T_{A_\gamma} \geq j]}. \quad (23)$$

Kun tiedetään (23), ja oletetaan että T_{A_γ} on riippumaton sarjasta $T_{A_\gamma}^{(1)}, T_{A_\gamma}^{(2)}, \dots$ saamme

$$\begin{aligned}
\mathbb{E}_v[Q_{J_v} - v] &= \mathbb{E}_v[\mathbb{E}_v[Q_{J_v} - v | Q_{J_{v-1}}]] \\
&= \sum_{k=1}^v \mathbb{E}_k[T_{A_\gamma} - k | v - Q_{J_{v-1}} = k, T_{A_\gamma} \geq k] \mathbb{P}_\infty[v - Q_{J_{v-1}} = k] \\
&= \sum_{k=1}^v \mathbb{E}_k[T_{A_\gamma} - k | T_{A_\gamma} \geq k] \mathbb{P}_\infty[v - Q_{J_{v-1}} = k] \\
&\xrightarrow{v \rightarrow \infty} \frac{\sum_{k=1}^{\infty} \mathbb{E}_k[T_{A_\gamma} - k | T_{A_\gamma} \geq k] \mathbb{P}_\infty[T_{A_\gamma} \geq k]}{\sum_{j=1}^{\infty} \mathbb{P}_\infty[T_{A_\gamma} \geq j]} \\
&= \frac{\sum_{k=1}^{\infty} \mathbb{E}_\infty[T_{A_\gamma} - k; T_{A_\gamma} \geq k]}{\mathbb{E}_\infty[T_{A_\gamma}]} \\
&= \frac{\sum_{k=1}^{\infty} \mathbb{E}_k[(T_{A_\gamma})^+]}{\gamma},
\end{aligned}$$

jolloin kohta (i) on todistettu. Seuraavaksi todistetaan kohta (ii). Kuten kohdassa (i), myös

$$\lim_{v \rightarrow \infty} \mathbb{E}_v[W_{K_v} - v] = \frac{\sum_{k=1}^{\infty} \mathbb{E}_k[(W - k)^+]}{\gamma}.$$

Kun tämän lisäksi sovelletaan lausetta (2.2), kohta (ii) on todistettu. Kohdan (iii) todistusta ei esitetä tässä, mutta se on todistettu artikkelissa [4]. \square .

Lauseen 2.3 kohta (iii) johtaa siihen että lause 2.2 on voimassa kaikilla $T \in \Delta_\gamma[4]$.

2.2 Kynnysarvo

Shiryaev-Robertsin menetelmässä kynnysarvo A on tärkeässä roolissa, jotta menetelmä osataan pysäyttää ajoissa. Aiemmissa luvuissa kynnysarvo on määritelty ehdon (6) mukaisesti, eli väärän hälytyksen havaitsemisajankohdan odotusarvon avulla. Tällöin arvo $A = A_\gamma$ valitaan siten, että $\mathbb{E}_\infty[T_{A_\gamma}] = \gamma$. Tällä tavoin laskettaessa tulee tietää etukäteen miten havainnot ovat jakautuneet muutoksen jälkeen. Tosielämässä kuitenkin ollaan yleensä tilanteessa, jossa ei tiedetä mikä tiheysfunktio tulee muutoksen jälkeen olemaan.

Kynnysarvo A voidaan kuitenkin määrittää toisellakin tavalla. Jos muutoksen jälkeistä tiheysfunktioita ei tiedetä, voidaan kynnysarvo laskea väärän hälytyksen todennäköisyyden p avulla. Tällöin kynnysarvo A valitaan niin, että

$$\mathbb{P} \left[\max_{1 \leq i \leq n} S_i(1) \geq A \right] = p, \tag{24}$$

jossa $(S_i(1))_{1 \leq i \leq n}$ on Shiryaev-Robertsin havaintolukusarja ehdolla, että ensimmäisten n havaintojen aikana muutosta ei tapahdu. Todennäköisyysarvo p on parametri, jonka menetelmän käyttäjä päättää. Väärän hälytyksen todennäköisyys tarkoittaa todennäköisyyttä, että ajanhetki T on sellainen, että hälytys tapahtuu vaikka havaintojen tiheysfunktiossa ei ole tapahtunut muutosta.

Menetelmän herkkyyttä sarjan S_n satunnaiseen vaihteluun voidaan säädellä todennäköisyyden p avulla. Kun väärän hälytyksen todennäköisyys p on matala, voidaan tarkastella pidemmän aikavälin muutoksia, sillä menetelmä ei aiheuta hälytystä herkästi sarjan S_n mahdollisesti satunnaisesta vaihtelusta, vaan tarvitaan vaihtelua useamman havainnon kohdalla jotta hälytys tapahtuu. Vastaavasti kun p on korkea, voidaan tarkastella lyhyemmän aikavälin muutoksia.[1]

Tässä työssä jatkossa kynnysarvon A laskemiseen hyödynnetään jälkimmäiseksi esiteltyä tapaa, eli A lasketaan käyttäen väärän hälytyksen todennäköisyyttä p . Käytännössä tämä tapahtuu niin, että A estimoidaan simuloimalla menetelmää kun muutosta ei tapahdu. Simuloinnissa luodaan siis havaintoja joiden jakauma on samanlainen kuin tarkasteltavien havaintojen jakauma ennen muutosta. Näin luoduille havainnoille käytetään Shiryaev-Robertsin menetelmää ja lasketaan millä kynnysarvolla A saadaan haluttu todennäköisyys p .

Tässä työssä simulointi tehdään luomalla tarkasteltavana olevan havaintosarjan muutosta edeltävän tilanteen mukaisesti jakautuneita havaintosarjoja 10 000 kertaa, jotta todennäköisyyden laskemiseen on riittävän monta havaintosarjaa, joille muutosta ei tapahdu. Tämä vaikuttaa paljon Shiryaev-Robertsin menetelmän tehokkuuteen, sillä simulointi tekee menetelmästä hitaan. Rinnakkaislaskennan avulla hitauteen voidaan hiukan vaikuttaa.

2.3 Poisson-jakauma

Tarkastellaan seuraavaksi menetelmää, kun havainnot (X_1, \dots, X_n) ovat Poisson-jakautuneet. Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus ja olkoon $x = (x_n)_{n \geq 1}$ sarja riippumattomia satunnaismuuttujia $X = (X_n)_{n \geq 1}$, joille pätee

$$X_n \sim \begin{cases} \text{Poisson}(\lambda), & \text{kun } 1 \leq n \leq v - 1, \\ \text{Poisson}(\lambda\rho_0), & \text{kun } v \leq n, \end{cases}$$

missä $\lambda > 0$ on tunnettu ja $\rho_0 > 0$ sekä $v \in \mathbb{N} \cup \{\infty\}$ ovat tuntemattomia. Nyt v on ajanhetki jolloin muutos tapahtuu ja ρ_0 on kerroin, jonka verran intensiteetti muuttuu. Merkitään nyt Shiryaev-Robertsin havaintolukusarjaa seuraavasti; $S(\rho_0) = (S_n(\rho_0))_{n \geq 1}$. Havaintoluku on määritelty kaikille $n \geq 1$

seuraavasti

$$S_n(\rho_0) = \sum_{k=1}^n \prod_{i=k}^n \frac{f_{\rho_0}(X_i)}{f_1(X_i)} = (1 + S_{n-1}(\rho_0)) \frac{f_{\rho_0}(X_n)}{f_1(X_n)},$$

missä f_{ρ_0} on Poisson-jakautuneen satunnaismuuttujan tiheysfunktio intensiteetillä $\lambda\rho_0$, kaikille $\rho_0 > 0$;

$$f_{\rho_0}(x) = e^{-\lambda\rho_0} \frac{(\lambda\rho_0)^x}{x!}, \quad x \in \mathbb{N}.$$

Havaintoluku $S_n(\rho_0)$ voidaan nyt kirjoittaa myös muodossa

$$S_n(\rho_0) = \sum_{k=1}^n e^{\lambda(\rho_0-1)(n-k+1) + \log(\rho_0) \sum_{i=k}^n X_i}.$$

Havaintolukuja $S_n(\rho_0)$ lasketaan niin kauan kunnes saavutetaan kynnyсарvo $A = s_n^*$. Havaitsemisajankohta $T(s_n^*)$ joka toteuttaa yhtälön (2), on ensimmäinen ajanhetki jolloin menetelmä havaitsee muutoksen. Kynnyсарvo s_n^* lasketaan väärän hälytyksen todennäköisyyden p avulla, kaavan (24) mukaisesti.

Kun $\lambda > 0$ ja $n \geq 0$ ovat tunnettuja ja $v \geq 1$ on tuntematon, ehdotetaan artikkelissa [1] intensiteetin suhdeluvun ρ_0 estimaatiksi lukua $\hat{\rho}_S(n)$, joka on määritelty seuraavasti:

$$\hat{\rho}_S(n) := \arg \max_{\rho_0 > 0} S_n(\rho_0).$$

Huomataan että estimaatti $\hat{\rho}_S(n)$ ei riipu muutosajankohdasta v . Täten jakauman intensiteetin muutoksen havaitsemista varten Shiryaev-Robertsin menetelmällä ei tarvitse tietää, milloin muutos tapahtuu. Muutoksen havaitsemista varten käytetään sarjaa $\tilde{S}_n = (\tilde{S}_{i,n})_{1 \leq i \leq n}$, joka on määritelty seuraavasti:

$$\tilde{S}_{i,n}(\hat{\rho}_S(n)) = \sum_{k=1}^i \prod_{j=k}^i \frac{f_{\hat{\rho}_S(n)}(X_j)}{f_1(X_j)}.$$

Arvoja $\tilde{S}_{i,n}$ lasketaan niin kauan kun $\tilde{S}_{i,n} \leq \tilde{s}_n^*$. Tästä seuraa, että jokaisen uuden havainnon kohdalla (kun n kasvaa) sarja \tilde{S}_n lasketaan aina uudelleen. Tällöin heti kun hälytys tapahtuu, menetelmä tarjoaa estimaatin muutoksen jälkeiselle intensiteetille.

2.4 Estimoidun intensiteetin muutoskertoimen konsistenssi

Tässä kappaleessa tarkastellaan intensiteetin muutoskertoimen ρ_0 estimaatin $\hat{\rho}_S(n)$ konsistenssia, kun sarja $(X_n)_{n \geq 1}$ on määritelty kuten luvussa 2.3

Määritelmä 2.4. Arvon ρ_0 estimaatin $\hat{\rho}(n)$ sanotaan olevan konsistentti jos $\hat{\rho}(n) \xrightarrow[n \rightarrow \infty]{P} \rho_0$.

Olkoon kaikille $1 \leq i \leq n$

$$\hat{\rho}_{i,n} = \frac{1}{\lambda(n-i+1)} \sum_{k=i}^n X_k$$

ja sarjan X_i, X_{i+1}, \dots, X_n suhdeluku

$$\begin{aligned} R_{i,n} &:= \prod_{k=i}^n \frac{f_\rho(X_k)}{f_1(X_k)} = e^{-\lambda(n-i+1)(\rho-1) + \log(\rho) \sum_{k=i}^n X_k} \\ &= e^{\lambda(n-i+1)(\hat{\rho}_{i,n} \log(\rho) - \rho + 1)}. \end{aligned}$$

Koska $\arg \max_{\rho > 0} R_{i,n}(\rho) = \arg \max_{\rho > 0} \log(R_{i,n}(\rho))$ jokaiselle $i \geq 1$, ja koska

$\frac{d}{d\rho} \log(R_{i,n}(\rho)) = 0$, kun $\rho = \frac{1}{\lambda(n-i+1)} \sum_{k=i}^n X_k$, niin tällöin funktio $\rho \mapsto R_{i,n}(\rho)$ saavuttaa maksimin pisteessä $\hat{\rho}_{i,n}$.

Huomioidaan myös, että sarjasta $(X_k)_{k \geq 1}$ ainoastaan äärellinen määrä alkioita poikkeaa sarjasta jossa alkiot ovat riippumattomia ja samoin jakautuneita. Täten suurten lukujen laki pätee, eli kaikille $i \geq 1$, $\hat{\rho}_{i,n}$ konvergoi melkein varmasti kohti muuttujaa ρ_0 ;

$$\lim_{n \rightarrow \infty} \hat{\rho}_{i,n} = \rho_0 \quad (\text{a.s.}) \quad (25)$$

Lause 2.5. Sarja $(\hat{\rho}_{i,n})_{n \geq 1}$ ei suppene tasaisesti kun $i \geq 1$.

Todistus. Tehdään vastaoletus: Sarja $(\hat{\rho}_{i,n})_{n \geq 1}$ suppenee tasaisesti kaikille $i \geq 1$.

Sarja $(\hat{\rho}_{i,n})_{n \geq 1}$ suppenee tasaisesti kaikille $i \geq 1$, jos on olemassa sellainen $p > 0$, että kaikille $\epsilon, \eta \geq 0$ on olemassa sellainen $N \geq 1$, että jokaiselle $n \geq N$ pätee $\mathbb{P}[\sup_{1 \leq i \leq n} |\hat{\rho}_{i,n} - p| > \epsilon] < \eta$.

Huomataan ensin, että $\sup_{1 \leq i \leq n} \hat{\rho}_{i,n} \geq \hat{\rho}_{n,n} := X_n$ kaikille $n \geq 1$. Koska sarjan $(X_n)_{n \geq 1}$ arvot ovat riippumattomia ja rajoittamattomia, X_n voi

saada minkä tahansa positiivisen kokonaislukuarvon aidosti positiivisella todennäköisyydellä. Määrätylle $p > 0$ ja mille tahansa $\epsilon > 0$ määritellään $A_n := \{X_n > p + \epsilon\}$ kaikille $n \geq 1$. Koska $\mathbb{P}[A_n]$ on aidosti positiivinen ja riippumaton arvosta n , voidaan löytää sellainen arvosta n riippumaton $\eta > 0$, jolle $\mathbb{P}[A_n] > \eta$. Nyt arvosta A_n saadaan $|X_n - p| > \epsilon$ ja $|\sup_{1 \leq i \leq n} \hat{\rho}_{i,n} - p| > \epsilon$.

Tämä on voimassa kaikille $n \geq 1$: Jokaiselle $p > 0$ voidaan löytää sellaiset $\epsilon, \eta > 0$, että kaikille $N \geq 1$ on olemassa vähintään yksi $n \geq N$ siten että $\mathbb{P}[\sup_{1 \leq i \leq n} \hat{\rho}_{i,n} - p| > \epsilon] > \eta$. Tämä on ristiriidassa vastaoletuksen kanssa. \square

Tarkastellaan seuraavaksi suhdelukujen $R_{i,n}(\rho) > 0$ summaa, eli havaintolukua (3)

$$S_n(\rho) = \sum_{i=1}^n R_{i,n}(\rho).$$

Estimaatti parametrille ρ_0 on

$$\hat{\rho}_S(n) = \arg \max_{\rho > 0} S_n(\rho).$$

Seuraavaksi tarkastellaan arvon $\hat{\rho}_S(n)$ arvoväliä.

Lemma 2.6. $\hat{\rho}_S(n) \in \left[\min_{1 \leq i \leq n} \hat{\rho}_{i,n}, \max_{1 \leq i \leq n} \hat{\rho}_{i,n} \right]$, kaikille $n \geq 1$.

Todistus. Koska jokaiselle $1 \leq i \leq n$, $\arg \max_{\rho > 0} R_{i,n} = \hat{\rho}_{i,n}$, kuvaus $\rho \mapsto R_{i,n}(\rho)$ on aidosti kasvava kun $\rho < \hat{\rho}_{i,n}$ ja aidosti vähenevä kun $\rho > \hat{\rho}_{i,n}$. Tästä seuraa että funktio $\rho \mapsto S_n(\rho)$ on myös aidosti kasvava kun $\rho < \min_{1 \leq i \leq n} \hat{\rho}_{i,n}$ ja aidosti vähenevä kun $\rho > \max_{1 \leq i \leq n} \hat{\rho}_{i,n}$. Lemma on seuraus tästä. \square

Seuraava lemma täsmentää arvon $\hat{\rho}_S(n)$ sijaintia.

Lemma 2.7. Mille tahansa $\epsilon > 0$ ja $i \geq 1$,

$$\lim_{n \rightarrow \infty} R_{i,n}(\rho_0 - \frac{\epsilon}{2}) - R_{i,n}(\rho_0 - \epsilon) = +\infty \quad (\text{a.s.}) \quad (26)$$

ja

$$\lim_{n \rightarrow \infty} R_{i,n}(\rho_0 + \frac{\epsilon}{2}) - R_{i,n}(\rho_0 + \epsilon) = +\infty \quad (\text{a.s.}) \quad (27)$$

Todistus. Todistetaan vain (27), sillä (26) saadaan symmetrian avulla, ja se löytyy myös artikkelista [1]. Sen lisäksi voidaan valita $i = 1$, sillä muut tapaukset ovat samanlaisia.

Tarkastellaan alkeistapausta $\omega \in \Omega$ jolle pätee yhtälö (25). Jätetään merkintä ω pois todistuksesta. Asetetaan $\epsilon > 0$ ja määritellään

$$\begin{aligned}x_n &:= \lambda n(\hat{\rho}_{1,n} \log \rho_1 - \rho_1 + 1), \\y_n &:= \lambda n(\hat{\rho}_{1,n} \log \rho_2 - \rho_2 + 1),\end{aligned}$$

missä $\rho_1 := \rho_0 + \frac{\epsilon}{2}$ ja $\rho_2 := \rho_0 + \epsilon$. Nyt $R_{1,n}(\rho_0 + \frac{\epsilon}{2}) - R_{1,n}(\rho_0 + \epsilon) = e^{x_n} - e^{y_n}$. Koska $e^{x_n} - e^{y_n} \geq x_n - y_n$, kun $x_n - y_n \geq 0$, lemma on todistettu kunhan näytetään että

$$x_n - y_n \xrightarrow[n \rightarrow \infty]{} +\infty.$$

Käytetään suppenemista $\hat{\rho}_{1,n} \xrightarrow[n \rightarrow \infty]{} \rho_0$ jotta löydetään riittävän suuri $N \geq 1$, jolle $\hat{\rho}_{1,n} > \rho_1 - \frac{\epsilon}{4}$, kun $n > N$. Näin suurelle n saadaan

$$\begin{aligned}x_n - y_n &= \lambda n(\hat{\rho}_{1,n} \log \rho_1 - \rho_1 + 1) - \lambda n(\hat{\rho}_{1,n} \log \rho_2 - \rho_2 + 1) \\&= \lambda n(\rho_2 - \rho_1 + \hat{\rho}_{1,n}(\log \rho_1 - \log \rho_2)) \\&> \lambda n\left(\rho_2 - \rho_1 + \rho_1(\log \rho_1 - \log \rho_2) - \frac{\epsilon}{4}(\log \rho_1 - \log \rho_2)\right) \\&= \lambda n\left((\rho_2 - \rho_1 \log \rho_2) - (\rho_1 - \rho_1 \log \rho_1) - \frac{\epsilon}{4}(\log \rho_1 - \log \rho_2)\right),\end{aligned}$$

missä $(\rho_2 - \rho_1 \log \rho_2) - (\rho_1 - \rho_1 \log \rho_1) > 0$, sillä $\rho \mapsto \rho - \rho' \log \rho$ on kasvava kun $\rho > \rho'$, ja $\log \rho_1 - \log \rho_2 < 0$. Täten $(\rho_2 - \rho_1 \log \rho_2) - (\rho_1 - \rho_1 \log \rho_1) - \frac{\epsilon}{4}(\log \rho_1 - \log \rho_2) > 0$, josta lemma seuraa. \square

Huomataan, että kun arvoa $\epsilon > 0$ vaihdellaan yhtälöissä (26) ja (27), saadaan lemmän sivutuloksena että kaikille $\rho > 0$ ja $i \geq 1$,

$$\lim_{n \rightarrow \infty} R_{i,n}(\rho) = +\infty \quad (\text{a.s.})$$

Lause 2.8. Ajanhetkelle $v = 1$, $\hat{\rho}_S(n) \xrightarrow[n \rightarrow \infty]{P} \rho_0$.

Todistus. Asetetaan $\epsilon, \eta > 0$. Lauseen todistamiseksi täytyy osoittaa, että on olemassa riittävän suuri $N \geq 1$, niin että jokaiselle $n \geq N$

$$\mathbb{P}[|\hat{\rho}_S(n) - \rho_0| \leq \epsilon] \geq 1 - 2\eta.$$

Aloitetaan suurten lukujen laista. Merkitään $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$, $n \geq 1$. Koska $(\bar{X}_n)_{n \geq 1}$ konvergoi melkein varmasti kohti lukua $\lambda \rho_0$, voidaan artikkelin [1] mukaan löytää sellainen $n_0 > 1$, jolle

$$\mathbb{P}[\forall n > n_0 \mid \frac{1}{\lambda} \bar{X}_n - \rho_0 \mid \leq \frac{\epsilon}{2}] \geq 1 - \frac{\eta}{3}.$$

Nyt määrätyle $n \geq n_0$ on voimassa

$$\mathbb{P}[\forall i \text{ s.t. } n - i \geq n_0, |\frac{1}{\lambda} \bar{X}_{n-i+1} - \rho_0| \leq \frac{\epsilon}{2}] \geq 1 - \frac{\eta}{3}. \quad (28)$$

Huomioidaan että sarjoille $(\frac{1}{\lambda} \bar{X}_1, \dots, \frac{1}{\lambda} \bar{X}_n)$ ja $(\hat{\rho}_{n,n}, \dots, \hat{\rho}_{1,n})$ on voimassa samat laskennalliset ominaisuudet. Merkitään nyt

$$A_n^1 := \{\forall i \leq n - n_0, |\hat{\rho}_{i,n} - \rho_0| \leq \frac{\epsilon}{2}\}. \quad (29)$$

Tällöin epäyhtälö (28) on ekvivalentti epäyhtälön

$$\mathbb{P}[A_n^1] \geq 1 - \frac{\eta}{3}$$

kanssa.

Olkoon nyt $L > 0$ sellainen, että

$$\mathbb{P}[\forall i \leq n_0, R_{i,n_0}(\hat{\rho}_{i,n_0}) \leq \frac{L}{n_0}] \geq 1 - \frac{\eta}{3}. \quad (30)$$

Koska kuvaus $\rho \mapsto R_{i,n_0}(\rho)$ saavuttaa maksiminsa kohdassa ρ_{i,n_0} kaikilla $i \geq 1$, epäyhtälöstä (30) seuraa, että

$$\mathbb{P}[\forall \rho > 0, \sum_{i=1}^{n_0} R_{i,n_0}(\rho) \leq L] \geq 1 - \frac{\eta}{3}.$$

Nyt kaikilla $n \geq n_0$ sarjoille $(R_{1,n_0}, \dots, R_{n_0,n_0})$ ja $(R_{n-n_0+1,n}, \dots, R_{n,n})$ on voimassa samat laskennalliset ominaisuudet. Tästä voimme nyt päätellä, että kaikille $n \geq n_0$,

$$\mathbb{P}[A_n^2] \geq 1 - \frac{\eta}{3}, \text{ kun } A_n^2 := \{\forall \rho > 0, \sum_{i=n-n_0+1}^n R_{i,n}(\rho) \leq L\}. \quad (31)$$

Kun L on kuten edellä, lemmän 2.7 perusteella voidaan löytää sellainen $n_1 \geq 1$, että kaikille $n \geq n_1$,

$$\mathbb{P}[A_n^3] \geq 1 - \frac{\eta}{3}, \text{ kun } A_n^3 := \{R_{1,n}(\rho_0 - \frac{\epsilon}{2}) - R_{1,n}(\rho_0 - \epsilon) \geq 2L\}. \quad (32)$$

Kun edellä saadut tulokset yhdistetään, saadaan että kaikille $n \geq \max\{n_0, n_1\}$, $\mathbb{P}[A_n] \geq 1 - \eta$, kun $A_n := A_n^1 \cap A_n^2 \cap A_n^3 \subset \Omega$.

Koska kuvaus $\rho \mapsto R_{i,n}(\rho)$ on kasvava kaikilla $\rho < \rho_0 - \epsilon$ ja $i \leq n - n_0$, epäyhtälöstä (29) seuraa että

$$R_{i,n}(\rho_0 - \frac{\epsilon}{2}) \geq R_{i,n}(\rho)$$

ja epäyhtälöstä (32) saadaan

$$R_{1,n}(\rho_0 - \frac{\epsilon}{2}) \geq R_{1,n}(\rho) + 2L.$$

Tällöin

$$\sum_{i=1}^{n-n_0} R_{i,n}(\rho_0 - \frac{\epsilon}{2}) \geq \sum_{i=1}^{n-n_0} R_{i,n}(\rho) + 2L. \quad (33)$$

Tästä tuloksesta seuraa yhdessä epäyhtälön (31) kanssa

$$\begin{aligned} \sum_{i=1}^n R_{i,n}(\rho_0 - \frac{\epsilon}{2}) &\geq \sum_{i=1}^n R_{i,n}(\rho) + 2L \\ &+ \left(\sum_{i=n-n_0+1}^n R_{i,n}(\rho_0 - \frac{\epsilon}{2}) - \sum_{i=n-n_0+1}^n R_{i,n}(\rho) \right) \\ &\geq \sum_{i=1}^n R_{i,n}(\rho) + L. \end{aligned} \quad (34)$$

Tämä tarkoittaa, että arvolla A_n , kaikille $\rho < \rho_0 - \epsilon$

$$S_n(\rho_0 - \frac{\epsilon}{2}) \geq S_n(\rho).$$

Tästä voidaan nyt päätellä, että kaikille $n \geq \max\{n_0, n_1\}$, $\mathbb{P}[\hat{\rho}_S(n) \geq \rho_0 - \epsilon] \geq 1 - \eta$. Symmetrian nojalla voidaan todeta että riittävän suurelle n , $\mathbb{P}[\hat{\rho}_S(n) \leq \rho_0 + \epsilon] \geq 1 - \eta$. Alkuperäinen väite seuraa, kun nämä kaksi tulosta yhdistetään. \square

Lemma 2.9. Mille tahansa äärelliselle $v > 1$, $\hat{\rho}_S(n) \xrightarrow[n \rightarrow \infty]{P} \rho_0$.

Todistus. Käyttäen samoja merkintöjä kuin edellä, jokaiselle $n > v$ Shiryayev-Robertsin havaintoluku voidaan kirjoittaa

$$S_n(\rho) = \sum_{i=1}^n R_{i,n}(\rho) = R_{v+1,n}(\rho) \sum_{i=1}^{v-1} R_{i,v-1}(\rho) + \sum_{i=v}^n R_{i,n}(\rho). \quad (35)$$

Merkitään $\hat{\rho} = \arg \max_{\rho > 0} R_{i,n}(\rho) = \frac{1}{\lambda(n-i+1)} \sum_{k=1}^n X_k$.

Suurten lukujen lain nojalla, kun $\epsilon, \eta > 0$, voidaan löytää sellainen $n_0 \geq 1$, että kaikille $n \geq n_0$,

$$\mathbb{P}[\tilde{A}_n^1] \geq 1 - \frac{\eta}{4}, \text{ kun } \tilde{A}_n^1 := \{\forall i \leq n - n_0, |n\hat{\rho}_{i,n} - \rho_0| \leq \frac{\eta}{2}\}. \quad (36)$$

Nyt koska sarjan $(X_n)_{n \geq v}$ arvot ovat riippumattomia ja samoin Poisson-jakautuneita parametrilla $\lambda \rho_0$, epäyhtälö (34) takaa että voidaan löytää sellaiset $n_1 > v$ ja $L > 0$, että kaikille $n > n_1$

$$\mathbb{P}[\tilde{A}_n^2] \geq 1 - \frac{\eta}{4}, \text{ kun } \tilde{A}_n^2 := \{\forall \rho < \rho_0 - \epsilon, \sum_{i=v}^n R_{i,n}(\rho_0 - \frac{\epsilon}{2}) \geq \sum_{i=v}^n n R_{i,n}(\rho) + L\}. \quad (37)$$

Nyt kun $n \geq n_0$, olkoon ρ_n^- sellainen että $\mathbb{P}[\rho_n^- \leq \min_{1 \leq i \leq n} \hat{\rho}_{i,n}] \geq 1 - \frac{\eta}{8}$ ja $0 < k < K$ sellainen että $\mathbb{P}[\forall \rho \in [\rho_n^-, \rho_n^- - \epsilon), k < \sum_{i=1}^{v-1} R_{i,v-1}(\rho) < K] \geq 1 - \frac{\eta}{8}$. Merkitään

$$\tilde{A}_n^3 := \{\rho_n^- \leq \min_{1 \leq i \leq n} \hat{\rho}_{i,n} \text{ ja } \forall \rho \in [\rho_n^-, \rho_0 - \epsilon), k < \sum_{i=1}^{v-1} R_{i,v-1}(\rho) < K\},$$

ja saadaan

$$\mathbb{P}[\tilde{A}_n^3] \geq 1 - \frac{\eta}{4}. \quad (38)$$

Nyt lemmasta 2.7 ja sen jälkeen tehdystä huomiosta seuraa, että voidaan löytää sellainen $n_2 \geq 1$, että kaikille $n \geq n_2$,

$$\mathbb{P}[\tilde{A}_n^4] \geq 1 - \frac{\eta}{4}, \text{ kun } \tilde{A}_n^4 := \left\{ \frac{R_{v,n}(\rho_0 - \frac{\epsilon}{2})}{R_{v,n}(\rho_0 - \epsilon)} \geq \frac{K}{k} \right\}. \quad (39)$$

Kun nämä tulokset yhdistetään ja kun $n \geq \max\{n_0, n_1, n_2\}$ määritetään $\tilde{A}_n := \tilde{A}_n^1 \cap \tilde{A}_n^2 \cap \tilde{A}_n^3 \cap \tilde{A}_n^4 \subset \Omega$. Tällöin $\mathbb{P}[\tilde{A}_n] \geq 1 - \eta$.

Koska kuvaus $\rho \mapsto R_{i,n}(\rho)$ on kasvava, niin epäyhtälöiden (36) ja (39) kanssa siitä seuraa että arvolla \tilde{A}_n , kaikille $\rho < \rho_0 - \epsilon$

$$R_{v,n}(\rho_0 - \frac{\epsilon}{2}) \geq \frac{K}{k} R_{v,n}(\rho).$$

Kun tämä tulos yhdistetään epäyhtälön (38) kanssa, saadaan kaikille $\rho \in [\rho_n^-, \rho_0 - \epsilon)$,

$$\begin{aligned} R_{v,n}(\rho_0 - \frac{\epsilon}{2}) \sum_{i=1}^{v-1} R_{i,v}(\rho_0 - \frac{\epsilon}{2}) \\ &\geq R_{v,n}(\rho) \frac{K \sum_{i=1}^{v-1} -i = 1^{v-1} R_{i,v-1}(\rho_0 - \frac{\epsilon}{2})}{k \sum_{i=1}^{v-1} -i = 1^{v-1} R_{i,v-1}(\rho)} \sum_{i=1}^{v-1} R_{i,v-1}(\rho) \\ &\geq R_{v,n}(\rho) \sum_{i=1}^{v-1} R_{i,v-1}(\rho). \end{aligned}$$

Tämä tulos yhdessä epäyhtälöiden (35) ja (37) kanssa antaa kaikille $\rho \in [\rho_n^-, \rho_0 - \epsilon)$, tuloksen

$$S_n(\rho_0 - \frac{\epsilon}{2}) > S_n(\rho).$$

Näin ollen arvolle $\tilde{A}_n, \tilde{\rho}_S(n) \notin [\rho_n^-, \rho_0 - \epsilon)$. Arvon ρ_n^- valinnasta seuraa, että arvolla $\tilde{A}_n, \tilde{\rho}_S(n) \geq \rho_0 - \epsilon$. Tämä voidaan päätellä samoin kuin lemmän 2.7 todistuksessa. \square

3 Muutoksen havaitseminen Shiryaev-Robertsin menetelmällä

Tässä kappaleessa esitellään kaksi algoritmia Shiryaev-Robertsin menetelmästä, kun aineisto on sarja riippumattomia Poisson-jakautuneita satunnaismuuttujia. Ensimmäistä algoritmia käytetään kun halutaan havaita muutos tasossa, eli jos tiheysfunktio muuttuu. Nimitystä taso tulee siitä, että kun aineisto on samoin jakautunut, muutos näyttää siltä kuin havainnot laskisivat tai nousisivat eri tasolle. Kuvasta 1 voidaan nähdä tällainen hyppy alaspäin. Toista algoritmia käytetään kun halutaan havaita muutos trendissä. Trendimuutoksen kohdalla tutkitaan muutoksen havaitsemista sarjasta, jossa riippumattomilla Poisson-jakautuneilla satunnaismuuttujilla on tasaisella vauhdilla muuttuva intensiteetti. Trendimuutos tarkoittaa siis muutosta intensiteetin muutuskertoimessa.

3.1 Shiryaev-Robertsin menetelmä tason muutokselle

Seuraavaksi esitetään vaiheittain Shiryaev-Robertsin menetelmän kulku kun pyritään havaitsemaan muutos jakauman tiheysfunktiossa.

Algoritmi 3.1 (Shiryaev-Robertsin menetelmä tason muutokselle). Menetelmässä on viisi askelta:

Askel 1. Aseta aika-askel $n = 1$;

Askel 2. Estimoi $\hat{\rho}_S(n)$;

Askel 3. Laske Shiryaev-Robertsin sarja $(\tilde{S}_{i,n}(\hat{\rho}_S(n)))_{1 \leq i \leq n}$;

Askel 4. Laske kynnyisarvo s_n^* ;

Askel 5. Jos $(\tilde{S}_{i,n}(\hat{\rho}_S(n)))_{1 \leq i \leq n} \geq s_n^*$, tee hälytys ja lopeta. Muussa tapauksessa aseta $n = n + 1$ ja siirry askeleeseen 2.

Käydään nyt algoritmin toiminta läpi esimerkin avulla.

Esimerkki 3.2. Olkoon havainnot

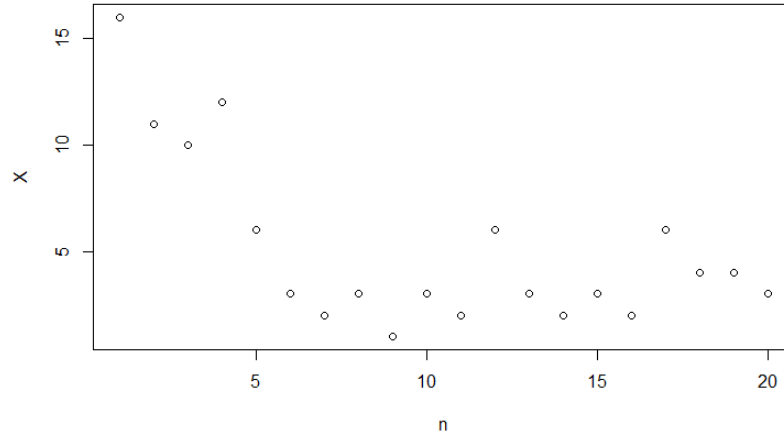
$$(16, 11, 10, 12, 6, 3, 2, 3, 1, 3, 2, 6, 3, 2, 3, 2, 6, 4, 4, 3)$$

sarja Poisson-jakaumasta simuloituja satunnaismuuttujia, joille pätee

$$X_n \sim \begin{cases} \text{Poisson}(12), & \text{kun } 1 \leq n \leq v - 1, \\ \text{Poisson}(12\rho_0), & \text{kun } v \leq n \leq 20. \end{cases}$$

n	$\hat{\rho}_S(n)$	$\tilde{S}_{n,n}$	s_n^*	Tila
1	1.333	1.827	52.642	Ei hälytystä
2	1.066	2.067	52.665	Ei hälytystä
3	0.932	3.138	80.848	Ei hälytystä
4	0.936	4.075	126.616	Ei hälytystä
5	0.764	11.070	126.861	Ei hälytystä
6	0.367	582.989	172.528	Hälytys!

Taulukko 1: Menetelmän kulku esimerkin 3.2 tapauksessa.



Kuva 1: Esimerkin 3.2 havainnot.

Ajankohta v jolloin tiheysfunktio muuttuu, sekä arvo ρ_0 ovat tietokoneen ar-
pomia siten, että $1 \leq v \leq 20$ ja $\rho \in (0, 2]$. Asetetaan $p = 0.01$ ja käytetään
algoritmia 3.1 tiheysfunktion muutoksen havaitsemiseen.

Taulukkoon 1 on listattu Shiryayev-Robertsin menetelmän tuottamia ar-
voja, joita tarkastelemalla havaitaan muutos havaintojen tiheysfunktiossa.
Muutos havaitaan ajanhetkellä $n = 6$, jolloin $\tilde{S}_{6,6} \geq s_6^*$. Taulukkoon on lis-
tattu sarjoista $(\tilde{S}_{i,n}(\hat{\rho}_S(n)))_{1 \leq i \leq n}$ vain viimeiset arvot, sillä rekursiokaavan
(4) perusteella sarjan viimeisin arvo on aina suurin. Kuvasta 1 voidaan myös
päättellä, että kun $n = 6$ arvot ovat selkeästi pienempiä.

3.2 Shiryaev-Robertsin menetelmä trendimuutokselle

Tässä kappaleessa tutkitaan muutoksen havaitsemista, kun sarjalla Poisson-jakautuneita satunnaismuuttujia on tasaisella vauhdilla muuttuva intensiteetti.

Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus ja $X = (X_n)_{n \geq 1}$ sarja riippumattomia satunnaismuuttujia, joille $X_n \sim \text{Poisson}(\lambda_n)$, $n \geq 1$ ja $\lambda_n := \alpha \lambda_{n-1}$ kaikille $1 \leq n \leq v-1$ ja $\lambda_n := \alpha' \lambda_{n-1}$ kaikille $v \leq n$, missä $1 \leq v \leq n$, $\lambda_0 > 0$, $\alpha > 0$, $\alpha' := \alpha \rho > 0$, $\rho > 0$ ja $\alpha \neq \alpha'$. Tässä työssä muuttujat λ_0 ja α ovat tunnettuja, muuttujia v ja α' ei tunneta. Muuttuja v merkitsee ajankohtaa jolloin muutos tapahtuu ja α' kertoo muutoksen jälkeen kuinka paljon havainnot kasvavat tai laskevat edelliseen havaintoon nähden.

Määrätylle $n \geq 1$ Shiryaev-Robertsin sarja $\xi = (\xi_{i,n})_{1 \leq i \leq n}$ kun $i \in \{1, \dots, n\}$ on:

$$\xi_{i,n}(\theta_n) := \sum_{k=1}^i \prod_{j=k}^i \frac{f_{\theta_n}^j}{f_{\theta_0}^j}(X_j), \quad (40)$$

jossa $f_{\theta_n}^i$ on Poisson-jakautuneen satunnaismuuttujan tiheysfunktio intensiteetillä

$$\lambda_i := \begin{cases} \alpha \lambda_{i-1}, & \text{jos } 1 \leq i \leq v-1 \\ \alpha \rho \lambda_{i-1}, & \text{jos } v \leq i \leq n \end{cases} \quad (41)$$

ja $\theta_n = (\rho, v)$. Havaintoluku (40) voidaan esittää myös rekursiivisesti seuraavasti:

$$\xi_{i,n} = (1 + \xi_{i-1,n}) \frac{f_{\theta_n}^n(X_i)}{f_{\theta_0}^n(X_i)}, \quad \xi_{0,n} = 0. \quad (42)$$

Aiemmin tarkasteltiin tason vaihtelua ja huomattiin, että vaihtelun havaitsemiseen ei muuttujan ρ estimointia varten tarvitse tietää ajankohtaa, jolloin taso vaihtuu. Trendinmuutoksen havaitsemismenetelmässä tämä ei päde. Menetelmä käyttää estimaatteja muuttujalle v ja muuttujalle ρ . Merkitään estimoitua parametriparia $\hat{\theta}_n := (\hat{\rho}_\xi(n), \hat{v}_\xi(n))$, missä $\hat{\rho}_\xi(n)$ on estimaatti kertoimesta ρ ja $\hat{v}_\xi(n)$ on estimaatti ajankohdasta v .

Parametri $\hat{\theta}_n$ voidaan käytännössä estimoida suurimman uskottavuuden menetelmällä, mutta joskus voi olla hyödyllisempää käyttää neliövirheen minimointia, etenkin jos trendi muuttuu vahvasti vähenevään suuntaan[1].

Parametri $\hat{\theta}_n$ on siis parametripari joka maksimoi havaintoluvun $\xi_{n,n}$:

$$\hat{\theta}_n = \arg \max_{\rho > 0, v > 0} \xi_{n,n}(\theta_n).$$

Satunnaismuuttujien ollessa Poisson -jakautuneita, voidaan arvo (40) kirjoittaa myös seuraavalla tavalla:

$$\xi_{i,n} = \sum_{k=1}^i \prod_{j=k}^i e^{-\lambda \alpha^j (\hat{\rho}_\xi(n))^{j-\hat{v}_\xi(n)-1}} \hat{\rho}_\xi(n)^{X_j} \mathbb{1}\{j \geq \hat{v}_\xi(n)\} + \mathbb{1}\{j < \hat{v}_\xi(n)\}.$$

Menetelmän mukaan arvoja $\xi_{i,n}$ lasketaan niin kauan kun $\xi_{i,n} \leq x_n^*$, missä x_n^* on kynnsarvo. Raja-arvo x_n^* valitaan yleensä väärän hälytyksen rajoituksen mukaan. Tämä rajoitus määritellään väärän hälytyksen todennäköisyyden p ja kynnsarvon x_n^* mukaan

$$\mathbb{P} \left[\max_{1 \leq i \leq n} \xi_{i,n}(\theta_0) \geq x_n^* \right] = p, \quad (43)$$

jossa $\xi_{i,n}(\theta_0)$ on sellainen Shiryaev-Robertsin sarja, jossa trendimuutosta ei tapahdu ja p on menetelmän käyttäjän päätettävissä oleva todennäköisyys.

Algoritmi 3.3 (Shiryaev-Robertsin menetelmä trendin muutokselle). Menetelmässä on viisi askelta:

Askel 1. Aseta aika-askel $n = 1$;

Askel 2. Estimoi arvot $\hat{\rho}_\xi(n)$ ja $\hat{v}_\xi(n)$;

Askel 3. Laske Shiryaev-Robertsin sarja $(\xi_{i,n})_{1 \leq i \leq n}$;

Askel 4. Laske kynnsarvo x_n^* ;

Askel 5. Jos $(\xi_{i,n})_{1 \leq i \leq n} \geq x_n^*$, tee hälytys ja lopeta. Muussa tapauksessa aseta $n = n + 1$ ja siirry askeleeseen 2.

4 Työkyvyttömyyseläkkeelle siirtyneiden tarkastelu

Työkyvyttömyyseläkkeelle siirtyvien ihmisten lukumäärää halutaan voida ennustaa mahdollisimman tarkkaan. Tulevien vuosien ennusteita varten käytetään aiempien vuosien tapauksia. Shiryayev-Robertsin menetelmän avulla voidaan havaita, jos havainnoissa tapahtuu suurta muutosta kasvavaan tai laskevaan suuntaan. Tässä työssä on aluksi tarkasteltu lukumäärää jossa on huomioitu kaikenikäiset ja kaikkien sairaslukujen mukaan työkyvyttömyyseläkkeelle siirtyneet.

Työkyvyttömyyseläkkeelle vuosittain siirtyvien määrä on tässä oletettu Poisson-jakautuneeksi. Työkyvyttömyys johtuu yleensä sairaudesta tai vammasta jonka vuoksi henkilö ei kykene työskentelemään. Suurin osa väestöstä on kuitenkin työkykyisiä, jolloin työkyvyttömyyseläkkeelle siirtyminen on harvinainen tapahtuma. Lisäksi on oletettu että tapaukset ovat riippumattomia toisistaan.

Työkyvyttömyyseläkkeelle siirtyneiden lukumäärä on jaettu Suomen väkiluvulla, jolloin on saatu työkyvyttömyyseläkkeliiritysmisaste. Tämän jälkeen luvut on kerrottu sadalla, jolloin saadaan prosenttiluvut. Sen lisäksi luvut on skaalattu vielä arvolla k , prosenttimuodossa olleiden lukujen ollessa desimaalilukuja, ja suurimmassa osassa käsiteltäviä tapauksia alle yhden arvoisia. Skaalauskerroin k valitaan aineiston suuruusluokan mukaan, ja sen avulla voidaan säädellä myös menetelmän herkkyyttä. Jos esimerkiksi $k = 1000$, tarkastellaan prosentin tuhannesosissa tapahtuvia muutoksia. Poisson-jakautuneet muuttujat ovat diskreettejä, joten jos tarkastellaan muutoksia oletuksella että muuttujat ovat Poisson-jakautuneet, alle yhden kokoiset desimaalilukuarvoista ei huomattaisi muutoksia ellei niitä skaalattaisi suuremmiksi.

Suomen väkiluvut eri vuosina on saatu Tilastokeskuksen sivuilta Suomen virallisesta tilastosta [7]. Jokaisen käsiteltävän aineiston kohdalla on käytetty sitä vastaavaa väkilukua, esimerkiksi kun tarkastellaan alle 35-vuotiaiden työkyvyttömyyseläkettä, käytetään alle 35-vuotiaiden lukumäärää väestössä. Lisäksi väestöstä on huomioitu vain 18 vuotta täyttäneet, sillä alle 18-vuotias ei voi saada työkyvyttömyyseläkettä. Alle 35-vuotiaalla tarkoitetaan tässä siis 18-34 -vuotiaita.

Alkuennusteet on tehty ennustamalla aineiston perusteella aikaa ennen vuotta 1999, ja ennusteen avulla on laskettu λ_0 ja α . Ennusteessa käytetään R-ohjelmistolla funktioita `auto.arima()` ja `forecast()`. Näiden avulla voidaan ennustaa uusia havaintoja aikasarjasta, eli ajan mukaan nousevaan järjestykseen järjestetystä aineistosta, käyttäen parhaiten aineistoon sopivaa

ARIMA-mallia, eli integroitua autoregressiivisen liukuvan keskiarvon mallia. Ennustetta varten aikasarja käännetään ensin toisin päin, minkä jälkeen R-ohjelmiston edellä mainittujen funktioiden avulla ennustetaan aikasarjan havaintojen avulla miten sarja jatkuu halutun suuruisen ajanjakson verran. Kun ennuste on tehty, se lisätään aikasarjaan, jonka jälkeen aikasarja käännetään jälleen, jotta se olisi ajan mukaan nousevassa järjestyksessä. Jokaisen tarkasteltavan ryhmän kohdalla on ennusteen tekoon käytetty hyödyksi erimittaisia ajanjaksoja datasta, jotta alkuennusteesta on saatu realistinen. Jos esimerkiksi aineistosta voidaan nähdä aluksi maltillista nousua, jonka jälkeen havainnot lähtevät laskuun, käytetään ennusteeseen aikaa ennen laskukautta, jotta saadaan ennusteeseen aineistoa joka käyttäytyy kuten havainnot ennen trendin muutosta.

4.1 Aineisto

Työkyvyttömyyseläkkeelle siirtyneiden lukumäärän trendimuutoksen havaitsemisessa on käytetty Eläketurvakeskukselta saatua aineistoa. Aineistossa on listattu työeläkkeelle kuukausittain siirtyneitä vuosien 1999-2018 aikana. Henkilö voi saada myös kansaneläkettä, jos työeläkettä ei ole karttunut lainkaan tai sitä on karttunut hyvin vähän. Aineisto on ryhmitelty sukupuolen, eläkelajin, diagnoosin ja iän mukaan.

Eläkelajit tässä aineistossa on luokiteltu täysiin eläkkeisiin sekä osaeläkeisiin.

Diagnoosit on luokiteltu seuraavasti; mielenterveyden ja käyttäytymisen häiriöt, verenkierroelinten sairaudet, tuki- ja liikuntaelinten sekä sidekudoksen sairaudet, muut sairaudet. Diagnoosilla tarkoitetaan siis pääasiallista syytä, eli pääsairautta jonka takia henkilö saa eläkettä. Työkyvyttömyyden syynä voi joskus olla pääsairauden lisäksi jokin merkittävä lisäsairaus, mutta luokittelussa otetaan huomioon vain pääsairaus. Sairausdiagnoosit perustuvat kansainväliseen ICD-10 -tautiluokitukseen.

Iällä tarkoitetaan henkilön ikää eläkkeen alkaessa. Ikäryhmät ovat; 18-34 -vuotiaat, 35-44 -vuotiaat, 45-54 -vuotiaat, 55-59 -vuotiaat, 60-62 -vuotiaat ja yli 62 -vuotiaat.

4.2 Menetelmä käytännössä

Tässä kappaleessa käydään läpi Shiryaev-Robertsin menetelmä askel askeleelta. Aineistona on käytetty kaikkia työkyvyttömyyseläkkeelle siirtyneitä vuosina 1999-2018 ja menetelmää toistetaan kunnes ollaan käyty läpi koko

n	Vuosi	$\hat{\rho}_\xi(n)$	$\hat{v}_\xi(n)$	$\xi_{n,n}$	x_n^*	Tila
1	1999	2.000	2	2.000	2.011	Ei hälytystä
2	2000	1.006	2	4.000	4.022	Ei hälytystä
3	2001	0.972	3	6.001	6.034	Ei hälytystä
4	2002	0.998	1	8.000	8.047	Ei hälytystä
5	2003	0.999	1	10.000	10.052	Ei hälytystä
6	2004	0.973	6	12.003	12.070	Ei hälytystä
7	2005	0.975	6	14.013	14.079	Ei hälytystä
8	2006	0.935	7	16.107	16.103	Hälytys!

Taulukko 2: Menetelmän kulku ensimmäisellä kierroksella kun tarkastellaan kaikkia työkyvyttömyyseläkkeelle jääneitä.

aineisto. Näiden tulosten pohjalta on myös tehty kuva 3. Skaalauskerroin $k = 1000$.

Esimerkki 4.1. Tarkastellaan Shiryaev-Robertsin menetelmää kaikkien työkyvyttömyyseläkkeellesiirtymisasteeseen alkuarvoilla $p = 0.1\%$, $\lambda_0 = 0.448$ ja $\alpha = 1.050$. Arvo λ_0 on tässä ennen skaalausta olevassa muodossa.

1. kierros:

Taulukkoon 2 on listattu ensimmäisen kierroksen tulokset kolmen desimaalin tarkkuudella. Shiryaev-Robertsin sarjasta on raportoitu vain viimeinen arvo $\xi_{n,n}$, sillä sarjan lukuja verrataan kynnsarvoon x_n^* , ja sarjan viimeinen arvo on aina suurin koska arvo on yhtälön (42) mukainen summa. Hälytys tapahtuu jos $\tilde{\xi}_{n,n} \geq x_n^*$. Taulukosta nähdään hälytyksen tapahtuvan ajanhetkellä $n = 8$, eli vuonna 2006. Arvo $\hat{v}_\xi(8) = 7$ kertoo, että ajanhetkellä 8 havaittu muutos on tapahtunut ajanhetkellä 7 eli vuonna 2005. Viive muutoksen tapahtumisesta sen havaitsemiseen on siis yksi vuosi. Uusi ennuste tehdään arvon $\hat{\rho}_\xi(8)$ avulla. Alkuarvot ovat olleet $\lambda_0 = 0.448$ ja $\alpha = 1.050$, joten muutoksen jälkeen, kaavan (41) mukaan $\lambda_i = \alpha \hat{\rho}_\xi(8) \lambda_{i-1} = 1.05 \cdot 0.935 \cdot \lambda_{i-1} = 0.982 \cdot \lambda_{i-1}$, kun $7 \leq i \leq 8$.

Kun hälytys tapahtuu, tehdään uusi ennuste uusilla arvoilla. Tämän jälkeen voidaan menetelmä ajaa uudestaan, lähtien nyt alkuperäisen aineiston ajanhetkestä $n = 8$, eli vuodesta 2006 liikkeelle. Uusi aineisto koostuu siis alkuperäisestä aineistosta, mutta vuodet 1999-2005 jätetään pois. Alkuaskel $n = 1$ tarkoittaa nyt vuotta 2006. Uudet alkuarvot ovat $\lambda_0 = 0.601$ ja $\alpha = 0.982$.

2. kierros:

n	Vuosi	$\hat{\rho}_\xi(n)$	$\hat{v}_\xi(n)$	$\xi_{n,n}$	x_n^*	Tila
1	2006	2.000	2	2.000	2.011	Ei hälytystä
2	2007	1.058	2	4.004	4.023	Ei hälytystä
3	2008	1.041	2	6.013	6.036	Ei hälytystä
4	2009	1.012	1	8.008	8.044	Ei hälytystä
5	2010	1.004	1	10.002	10.060	Ei hälytystä
6	2011	1.002	1	12.001	12.074	Ei hälytystä
7	2012	0.900	7	14.038	14.092	Ei hälytystä
8	2013	0.964	6	16.068	16.089	Ei hälytystä
9	2014	0.944	7	18.180	18.109	Hälytys!

Taulukko 3: Menetelmän kulku toisella kierroksella.

Taulukkoon 3 on listattu toisen kierroksen tulokset. Nyt hälytys tapahtuu ajanhetkellä $n = 9$ eli vuonna 2014 ja muutos on tapahtunut ajanhetkellä $\hat{v}_\xi(9) = 7$ eli vuonna 2012. Nyt viive muutoksen tapahtumisesta sen havaitsemiseen on kaksi vuotta. Muutoksen jälkeen $\lambda_i = \alpha \hat{\rho}_\xi(9) \lambda_{i-1} = 0.982 \cdot 0.944 \cdot \lambda_{i-1} = 0.927 \cdot \lambda_{i-1}$, kun $7 \leq i \leq 9$. Uusilla arvoilla tehdään uusi ennuste, jonka jälkeen menetelmä ajetaan taas uudestaan lähtien liikkeelle edellisellä kerralla käytetyn aineiston ajanhetkestä $n = 9$ eli vuodesta 2014. Uudet alkuarvot ovat $\lambda_0 = 0.524$ ja $\alpha = 0.927$.

3. kierros:

Kolmannen kierroksen tulokset on listattu taulukkoon 4. Hälytys tapahtuu ajanhetkellä $n = 4$ eli vuonna 2017 ja muutos on tapahtunut ajanhetkellä $\hat{v}_\xi(4) = 2$ eli vuonna 2015. Viive muutoksen tapahtumisesta sen havaitsemiseen on kaksi vuotta. Muutoksen jälkeen $\lambda_i = \alpha \hat{\rho}_\xi(4) \lambda_{i-1} = 0.927 \cdot 1.068 \cdot \lambda_{i-1} = 0.990 \cdot \lambda_{i-1}$, kun $2 \leq i \leq 4$. Uusilla arvoilla tehdään uusi ennuste, jonka jälkeen menetelmä ajetaan taas uudestaan lähtien liikkeelle edellisellä kerralla käytetyn aineiston ajanhetkestä $n = 4$ eli vuodesta 2017. Uudet alkuarvot ovat $\lambda_0 = 0.421$ ja $\alpha = 0.99$. Neljännen kierroksen tulokset on listattu taulukkoon 5, josta nähdään että hälytystä ei tapahdu ollenkaan, joten uutta ennustetta ei tehdä ja menetelmä päättyy.

n	Vuosi	$\hat{\rho}_\xi(n)$	$\hat{v}_\xi(n)$	$\xi_{n,n}$	x_n^*	Tila
1	2014	2.000	2	2.000	2.012	Ei hälytystä
2	2015	1.061	2	4.003	4.022	Ei hälytystä
3	2016	1.070	2	6.026	6.037	Ei hälytystä
4	2017	1.068	2	8.084	8.045	Hälytys!

Taulukko 4: Menetelmän kulku kolmannella kierroksella.

n	Vuosi	$\hat{\rho}_\xi(n)$	$\hat{v}_\xi(n)$	$\xi_{n,n}$	x_n^*	Tila
1	2017	2.000	2	2.000	2.012	Ei hälytystä
2	2018	1.074	2	4.004	4.021	Ei hälytystä

Taulukko 5: Menetelmän kulku neljännellä kierroksella.

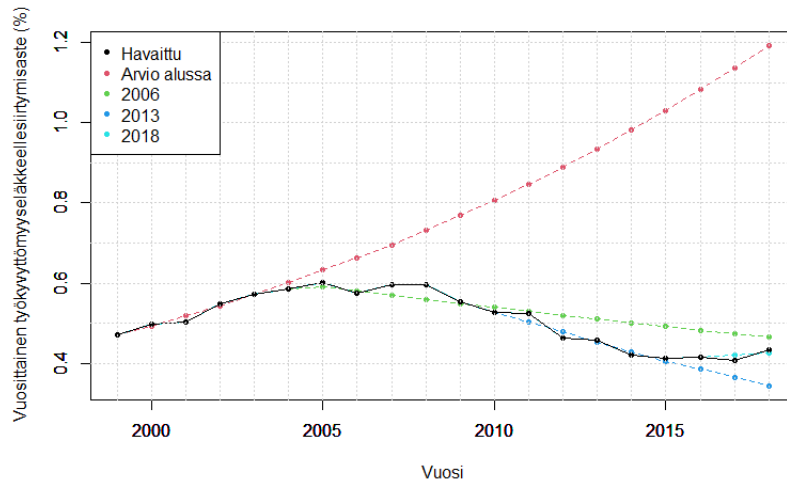
4.3 Väärän hälytyksen todennäköisyyden ja skaalauskerroimen vaikutus

Tässä kappaleessa tarkastellaan käytännössä miten Shiryaev-Robertsin menetelmän herkkyyttä voidaan säätää väärän hälytyksen todennäköisyyden avulla, sekä skaalauskerroimen avulla. Aineistona on käytetty kaikkia työvyökyttömyyseläkkeelle siirtyneitä vuosien 1999-2018 aikana. Aineistolle suoritetaan algoritmi 3.3, ja kun tapahtuu hälytys ja algoritmin suoritus lopetetaan, tehdään uusi ennuste. Tämän jälkeen korjataan intensiteetin arvo edellisen algoritmin antaman arvon $\hat{\rho}_\xi(n)$ avulla ja algoritmi suoritetaan uudestaan alkaen siitä ajanhetkestä mihin algoritmi viimeksi pysähtyi.

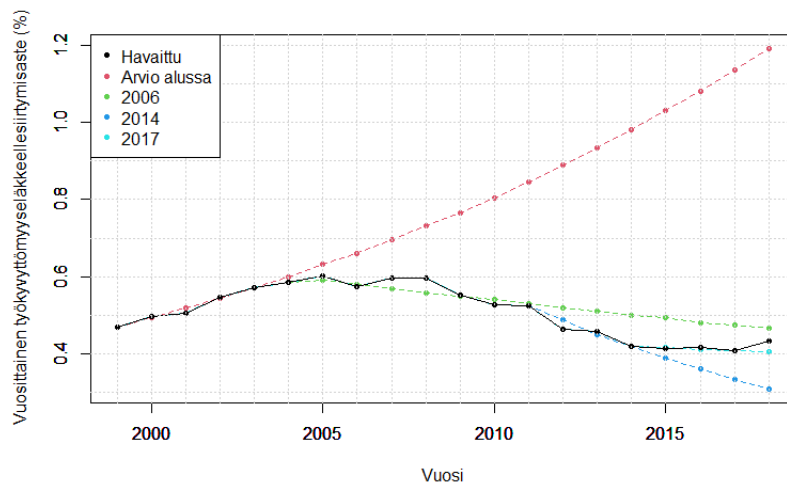
Tarkastellaan ensin väärän hälytyksen todennäköisyyden vaikutusta, kun skaalauskerroin on 1000. Kynnysarvot on laskettu käyttäen kolmea eri väärän hälytyksen todennäköisyyden arvoa p .

Kuvista 2, 3 ja 4 nähdään miten väärän hälytyksen todennäköisyyden valitseminen, ja siten kynnysarvo vaikuttaa trendimuutoksen havaitsemiseen. Kuvassa 2 väärän hälytyksen todennäköisyys on 1%, kuvassa 3 se on 0.1% ja kuvassa 4 se on 0.01%. Jokaisessa tapauksessa trendimuutos havaitaan kolme kertaa, mutta havainnot tapahtuvat eri ajankohtina. Esimerkiksi kuvissa 2 ja 3 ensimmäinen trendimuutos havaitaan vuonna 2006, kun taas kuvassa 4 se havaitaan vuonna 2007. Kuten kappaleessa 2.2 todettiin, menetelmän herkkyyttä muutoksille voidaan säätää väärän hälytyksen todennäköisyyden avulla. Kun todennäköisyys on pieni, muutos havaitaan myöhemmin.

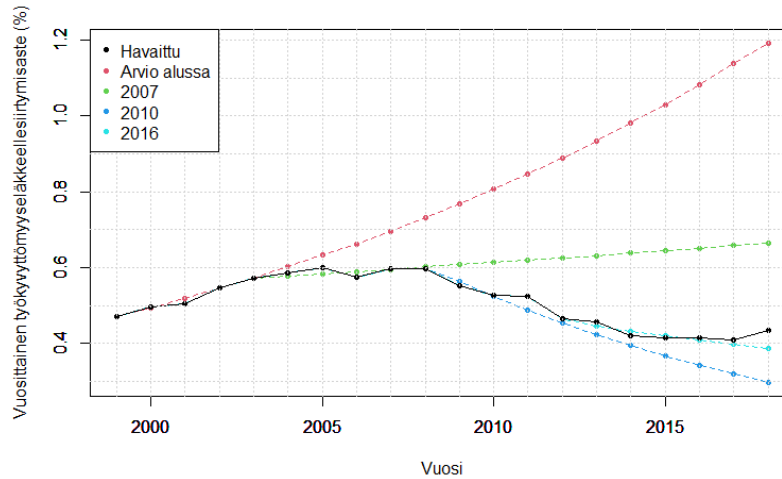
Kuvista voidaan nähdä trendin olevan aluksi nouseva, kunnes vuoden 2005 jälkeen tulee pieni notkahdus, mutta trendi on muuten melko tasai-



Kuva 2: Trendimuutoksen havaitseminen työkyvyttömyyseläkkeellesiirtymisasteesta vuosina 1999-2018, kun väärän hälytyksen todennäköisyys on 1%.



Kuva 3: Trendimuutoksen havaitseminen työkyvyttömyyseläkkeellesiirtymisasteesta vuosina 1999-2018, kun väärän hälytyksen todennäköisyys on 0.1%.

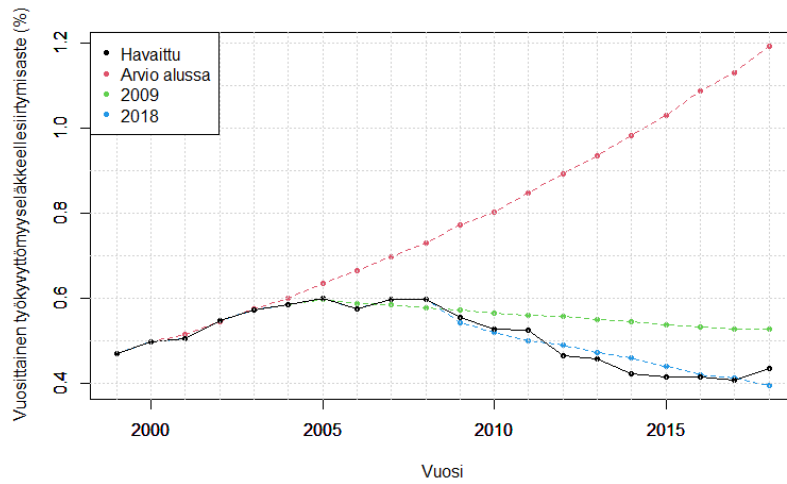


Kuva 4: Trendimuutoksen havaitseminen työkyvyttömyyseläkkeellesiirtymisasteesta vuosina 1999-2018, kun väärän hälytyksen todennäköisyys on 0.01%.

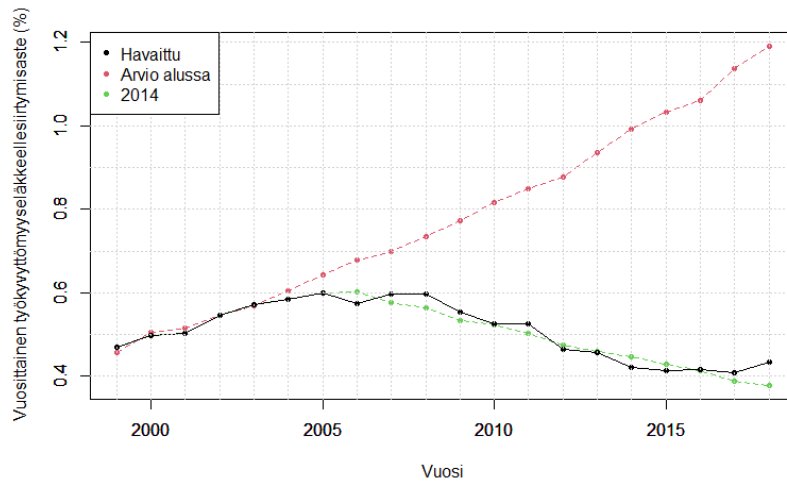
nen ja lähtee laskuun vuoden 2008 jälkeen. Kuvissa 2 ja 3 voidaan nähdä ensimmäisen hälytyksen aiheutuneen tästä notkahduksesta. Kun väärän hälytyksen todennäköisyys on ollut 0.01% ei vuoden 2005 jälkeen tapahtuva notkahdus aiheuta hälytystä. Lisäksi väärän hälytyksen todennäköisyyden ollessa 0.01%, havaitaan toinen muutos nopeammin, sillä trendimuutos on suurempi. Koska kuvassa 4 notkahdus ei aiheuttanut ennustetta jonka kulmakerroin on alle 1, kulmakertoimen muutos on suurempi kun havainnot lähtevät laskuun vuoden 2008 jälkeen. Toisaalta notkahduksen vuoksi tapahtunut trendimuutoksen havaitseminen kuvissa 2 ja 3 saa ennusteen lähtevän laskevaan suuntaan, kuten havainnoillekin käy hieman myöhemmin.

Eri väärän hälytyksen todennäköisyyden suuruusluokissa on siis erilaisia vaikutuksia. Mitä pienempi arvo p on, sitä myöhemmin muutokset havaitaan, mutta ne ovat tarkempia eivätkä aiheudu pienistä satunnaisista aineiston heilahteluista. Jos taas halutaan huomata mahdolliset muutokset nopeasti, on arvon p oltava suurempi, jolloin riskinä on väärin hälytysten kasvu.

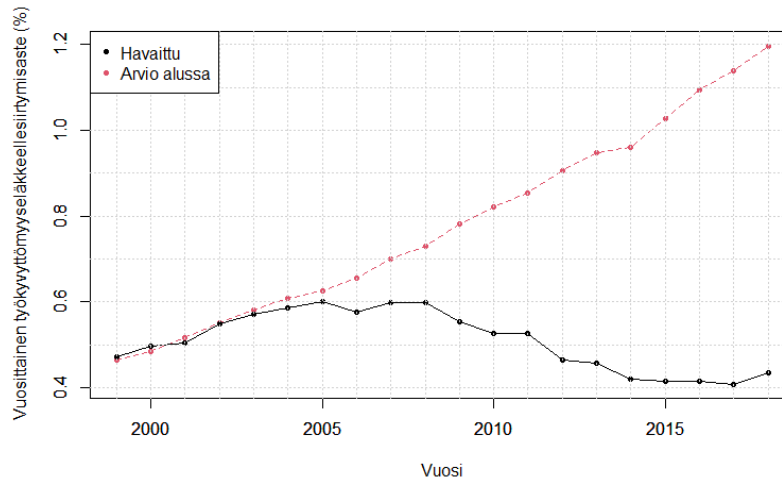
Skaalauskerroin vaikuttaa myös menetelmän herkkyyteen. Tarkastellaan seuraavaksi menetelmää eri skaalauskerroimilla, kun väärän hälytyksen todennäköisyys $p = 1\%$. Kuvassa 5 kerroin on 100, kuvassa 6 kerroin on 10 ja kuvassa 7 kerroin on 1. Voidaan huomata että mitä suurempi kerroin on, sitä herkempi menetelmä on havaitsemaan muutoksia. Skaalauskerroimen ol-



Kuva 5: Trendimuutoksen havaitseminen työkyvyttömyyseläkkeellesiirtymisasteesta vuosina 1999-2018, kun skaalauskerroin on 100.



Kuva 6: Trendimuutoksen havaitseminen työkyvyttömyyseläkkeellesiirtymisasteesta vuosina 1999-2018, kun skaalauskerroin on 10.



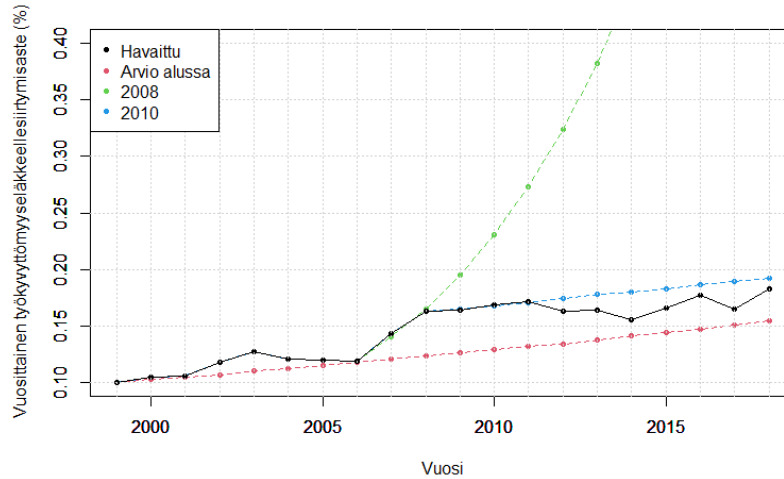
Kuva 7: Trendimuutoksen havaitseminen työkyvyttömyyseläkkeellesiirtymisasteesta vuosina 1999-2018, kun skaalauskerroin on 1.

lessa 1 muutoksia ei havaita ollenkaan. Tämä johtuu siitä, että simuloidut arvot joiden avulla lasketaan kynnsarvo, ovat samaa suuruusluokkaa kuin tarkasteltavat arvot.

4.4 Työkyvyttömyyseläkkeellesiirtymisasteen trendimuutos

Kun tarkastellaan kuvien 2, 3 ja 4 avulla työkyvyttömyyseläkkeellesiirtymisastetta kaikkien työkyvyttömyyseläkkeelle siirtyneiden osalta, voidaan huomata että trendi on ollut ensin melko tasaisessa nousussa, mutta lähtenyt laskuun vuoden 2008 jälkeen. Laskua selittävät monet tekijät, joista on kerrottu eläketurvakeskuksen tutkimuksessa[3]. Nykyään suurin osa eläkkeelle jäävistä siirtyy suoraan vanhuuseläkkeelle, kun taas 2000-luvun alussa monet olivat ensin työkyvyttömyys- tai työttömyyseläkkeellä ennen vanhuuseläkkeelle siirtymistä. Vuonna 2005 tehtiin uudistus, jonka mukaan 63-vuotta täyttäneiden uudet työkyvyttömyyseläkkeet muuttuivat vanhuuseläkkeiksi, kun taas ennen vuotta 2006 alkaneissa työkyvyttömyyseläkkeissä ikäraja on ollut 65-vuotta.

Tutkimuksessa kerrotaan myös ikääntyneiden työllistymisen parantumisesta, joka vähentää työkyvyttömyyseläkkeensaajien määrää. Lisäksi suuret ikäluokat ovat siirtyneet eläkkeensaajiksi, mikä on myös vähentänyt



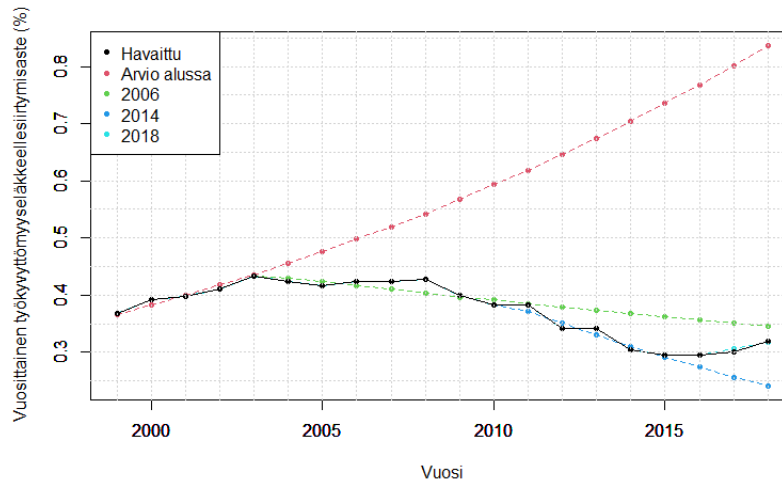
Kuva 8: Trendimuutoksen havaitseminen alle 35-vuotiaiden työkyvyttömyyseläkkeellesiirtymisasteesta.

työkyvyttömyyseläkkeelle siirtyvien lukumäärää.

4.5 Ikäryhmät

Tässä luvussa tarkastellaan trendimuutoksia eri ikäryhmien sisällä, kun väärän hälytyksen todennäköisyytenä on käytetty arvoa 0.1% ja skaalauskerroin $k = 1000$. Poikkeuksena 60-62 -vuotiaat, jossa väärän hälytyksen todennäköisyytenä on käytetty arvoa 0.01%, sillä tässä ikäryhmässä menetelmä havaitsi hyvin herkästi muutokset. Aloitetaan alle 35-vuotiaista.

Alle 35-vuotiaiden työkyvyttömyyseläkkeellesiirtymisaste on esitetty kuvassa 8. Alkuennusteen mukaan trendi olisi hieman nouseva, mutta vuonna 2008 havaitaan trendimuutos, jolloin ennuste lähtee jyrkkään nousuun. Vuonna 2005 tedyissä sääntömuutoksissa tapahtui muutoksia myös karttumissäännöissä, siten että pienikin työ kartuttaa eläkettä[3]. Tämän vaikutus koskee etenkin nuoria henkilöitä joiden työelämä ei ole ehtinyt päästä kunnolla alkuun ennen työkyvyttömyyden alkamista. Sääntömuutoksen vaikutus voidaan nähdä kuvasta 8 pienenä nousuna vuoden 2005 jälkeen, joka aiheuttaa myös trendimuutoksen. Vuoden 2008 jälkeen kasvu tasoittuu, jonka myötä vuonna 2010 havaitaan uusi trendimuutos ja tehdään uusi, lievästi nouseva ennuste.

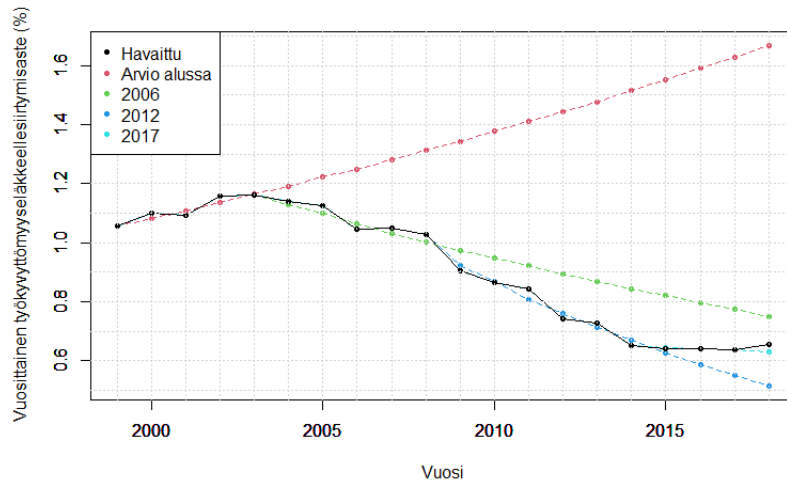


Kuva 9: Trendimuutoksen havaitseminen 35-44 -vuotiaiden työkyvyttömyyseläkkeellesiirtymisasteesta.

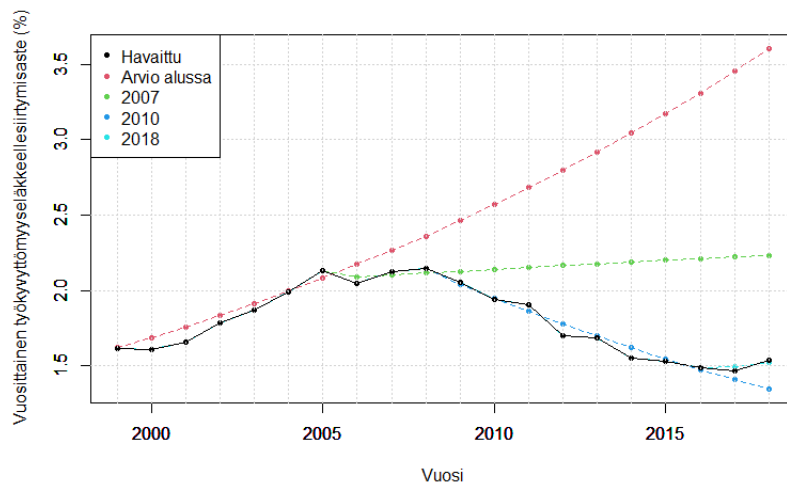
Kuvassa 9 nähdään miten 35-44 -vuotiaiden työkyvyttömyyseläkkeellesiirtymisaste on muuttunut. Alkuun trendi on hieman kasvavaa, kunnes se kääntyy laskuun. Ensimmäinen trendimuutos on havaittu vuonna 2006, jonka mukaan tehty ennuste lähtee loivasti laskuun, kun taas vuonna 2014 tehty ennuste lähtee hieman jyrkempään laskuun. Lopussa voidaan huomata pientä nousua, joka havaitaan trendimuutoksena vuonna 2018. Kuvio on hyvin samanlainen kuin mitä se oli tarkasteltaessa kaikkia ikäluokkia.

45-54 -vuotiaiden työkyvyttömyyseläkkeellesiirtymisaste on esitetty kuvassa 10. Tässä ikäluokassa trendi on aluksi kasvava, mutta kääntyy melko jyrkkään laskuun kun trendiä verrataan aluksi tehdyn arvion trendiin. Muutos havaitaan ja uusi ennuste tehdään vuonna 2006. Lisäksi vuonna 2012 havaitaan jälleen trendimuutos ja tehdään entistä jyrkempään laskuun lähtevä ennuste. Kuvasta voidaan kuitenkin huomata havaintojen tasoittuvan vuodesta 2014 eteenpäin, joka havaitaan myös uutena trendimuutoksena vuonna 2017.

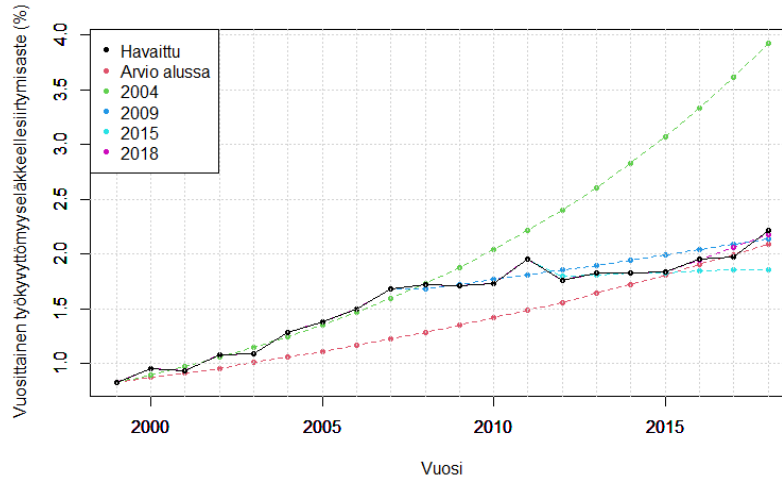
Myös 55-59 -vuotiaiden trendimuutos kääntyy laskuun, sen ollessa aluksi kasvava, kuva 11 näyttää tämän ikäryhmän työkyvyttömyyseläkkeellesiirtymisasteen. Vuonna 2007 havaittu trendimuutos aiheuttaa ennusteen joka on hieman alkuperäistä ennustetta loivemmin kasvava. Trendimuutoksen kääntyminen laskuun havaitaan vuonna 2010. Lopuksi huomataan trendimuutoksen tasoittuvan, ja havaintojen hieman kasvavan joka aiheuttaa vielä vuonna



Kuva 10: Trendimuutoksen havaitseminen 45-54 -vuotiaiden työkyvyttömyyseläkkeellesiirtymisasteesta.



Kuva 11: Trendimuutoksen havaitseminen 55-59 -vuotiaiden työkyvyttömyyseläkkeellesiirtymisasteesta.



Kuva 12: Trendimuutoksen havaitseminen 60-62 -vuotiaiden työkyvyttömyyseläkkeellesiirtymisasteesta.

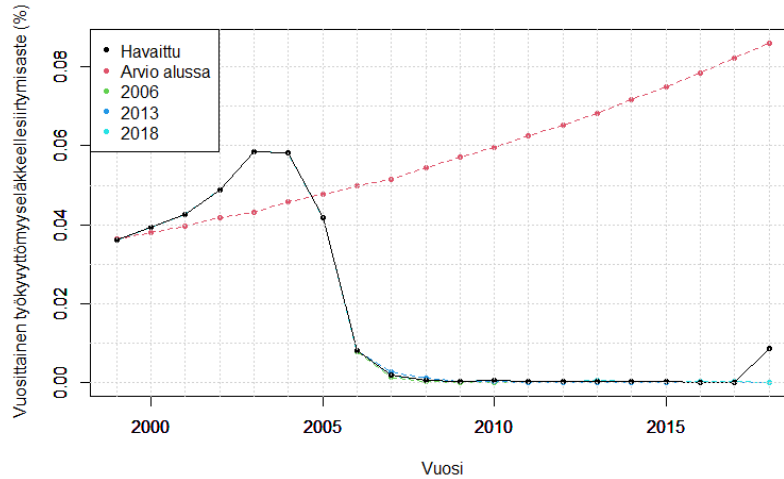
2018 trendimuutoksen.

Kuvassa 12 on esitetty 60-62 -vuotiaiden työkyvyttömyyseläkkeellesiirtymisaste, jossa voidaan huomata trendin olevan melkein koko tarkastelujakson ajan nouseva, mutta nousu hiipuu, kunnes vuonna 2018 tehdyn ennusteen mukaan trendimuutos lähtee jälleen kasvuun.

Kuvissa 2, 9, 11, 12 voidaan havaita työkyvyttömyyseläkkeellesiirtymisasteen nousua vuonna 2018. Blogikirjoituksessa [8] kerrotaan työkyvyttömyyseläkkeelle siirtyneiden määrän kääntyneen nousuun vuonna 2018. Kirjoituksessa mainitaan työkyvyttömyyseläkkeelle siirtyneiden määrän lisääntymisen mahdollisesti johtuvan suotuisasta talouskehityksestä. Kun työllisyystilanne on hyvä, myös heikomman terveydentilan omaavat valikoituvat töihin ja sairastamisen kynnyks on pienempi kun pelko työpaikan menettämisestä on pieni.

Aiemmin kerrottiin vuoden 2005 sääntömuutoksista ja miten ikäraja on muuttunut 65-vuotiaista 63-vuotiaisiin. Tämän muutoksen vaikutus nähdään hyvin selkeästi kuvasta 13, jossa on huomioitu vain yli 62-vuotiaat, jolloin luonnollisesti havainnot ovat lähellä nollaa pian vuoden 2005 jälkeen.

Kuvassa 14 nähdään yli 60-vuotiaiden työkyvyttömyyseläkkeellesiirtymisaste. Tässä kuvassa nähdään myös hyvin vuoden 2005 muutosten vaikutus, sillä trendi lähtee laskuun vuoden 2010 tehdyn ennusteen mukaan. Kuvassa voidaan huomata havaintojen lisääntyneen vuonna 2018. Blogikirjoi-

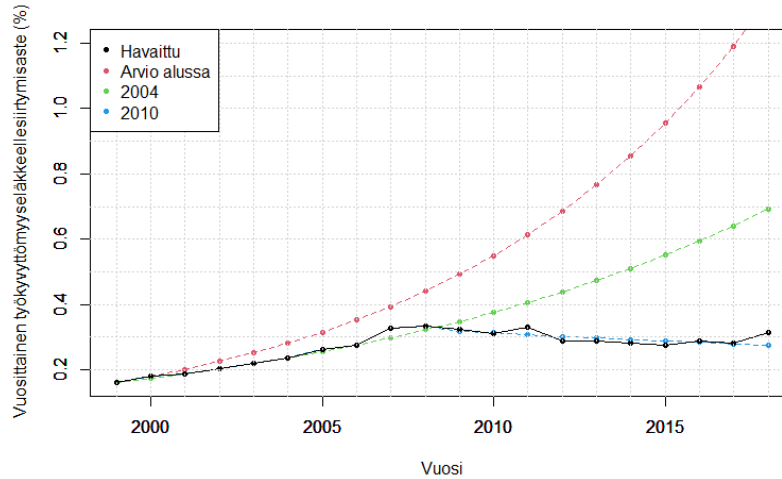


Kuva 13: Trendimuutoksen havaitseminen yli 62-vuotiaiden työkyvyttömyyseläkkeellesiirtymisasteesta.

tuksessa [8] kerrotaan vuoden 2017 eläkeuudistuksesta, joka nosti vanhuuseläkkeen alaikärajaa 63 vuodesta kolmella kuukaudella jokaista syntymävuosi-ikäluokkaa kohti, kunnes saavutetaan 65 vuoden alaikäraja. Tästä muutoksesta kerrotaan myös tutkimuksessa [3]. Koska työkyvyttömyyseläkkeelle siirtyminen on mahdollista oman ikäluokan vanhuuseläkkeen alarajaan saakka, on tämä muutos nostanut työkyvyttömyyseläkkeelle siirtymisen ikärajaa vuonna 2018. Tämä muutos voidaan nähdä havaintojen nousuna vuonna 2018. Shiryayev-Robertsin menetelmä ei kuitenkaan havainnut muutosta nousevaan suuntaan, mutta todennäköisesti näin tulisi tapahtumaan kun havaintoja saadaan lisää.

Kun kuvaa 13 verrataan kuvaan 12, voidaan huomata ikäluokkien kokoeuron vaikutus. Kuvassa 14 viimeisimmän ennusteen mukaan trendi on loivassa laskussa, kun taas kuvassa 12 voidaan havaita trendin olevan nouseva. Kuitenkin vuoden 2005 tehtyjen sääntömuutosten ja kuvan 13 perusteella vuoden 2007 jälkeen yli 62-vuotiailla ei alkanut työkyvyttömyyseläkkeitä juuri ollenkaan. Eläkkeelle siirtyneiden määrät kuvissa 14 ja 12 voidaan siis ajatella saman suuruisina. Trendin suunnan ero selittyy siis ikäluokkien kokoerolla.

Vuoden 2005 tehtyjen muutosten vuoksi tarkastellaan seuraavaksi työkyvyttömyyseläkkeellesiirtymisasteen trendimuutosta 18-62-vuotiaiden osalta. Näin saadaan parempi kuva nykyisten sääntöjen mukaisen työkyvyttömyyseläkkeellesiirtymisasteen trendimuutoksista. Kuvassa 15 nähdään työ-



Kuva 14: Trendimuutoksen havaitseminen yli 60-vuotiaiden työkyvyttömyyseläkkeellesiirtymisasteesta.

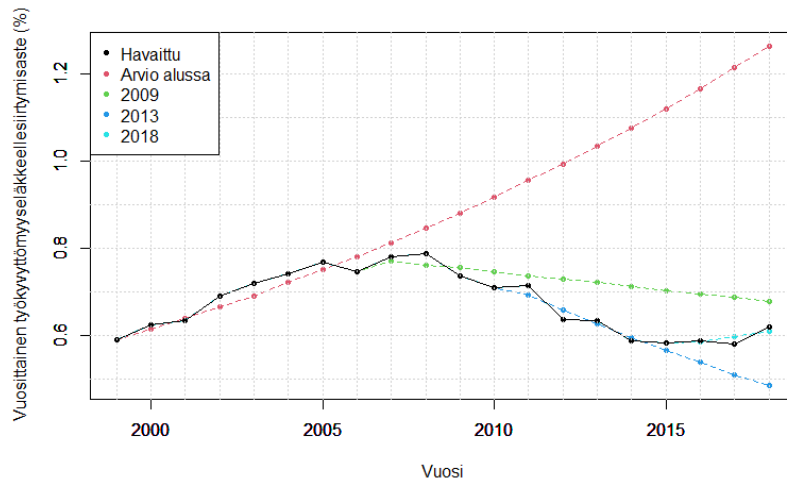
kyvyttömyyseläkkeellesiirtymisasteen trendimuutoksen havaitseminen 18-62-vuotiaiden osalta. Tässä kuvassa toistuu hyvin samankaltainen kuvio havaittujen arvojen osalta mitä kuvissa 2, 3 ja 4. Ensin työkyvyttömyyseläkkeellesiirtymisasteen trendi on nousevaa, mutta alkaa tasaantua vuoden 2005 jälkeen ja lähtee laskuun. Kuvassa 15 ei kuitenkaan havaita tasaantumista, vaan pelkästään trendin lähteminen laskuun joka havaitaan vuonna 2009, sekä 2013. Lopuksi trendi jälleen tasaantuu ja lähtee hieman kasvuun, joka huomataan vuoden 2018 ennusteesta.

Kun vertaillaan kuvia 3 ja 15, voidaan huomata että yksin yläikärajan muutos ei selitä trendin muutosta laskevaan suuntaan.

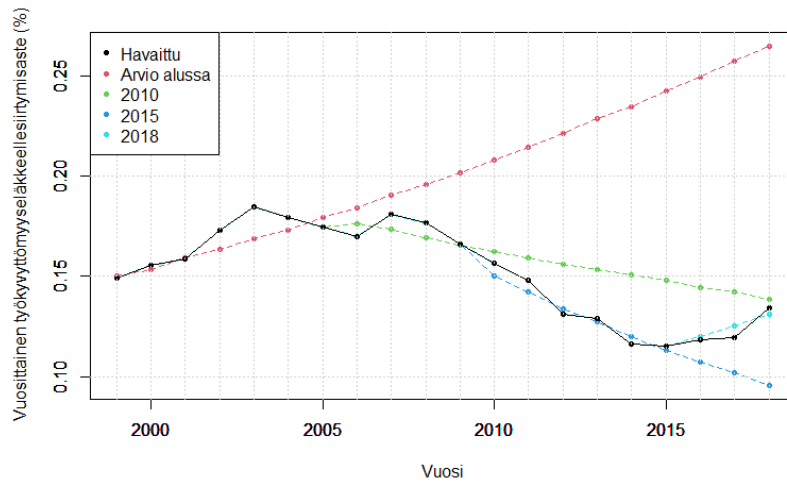
4.6 Sairauspääryhmät

Tässä luvussa tarkastellaan trendimuutoksia kahden eri pääsairausryhmän kohdalla; mielenterveyden ja käyttäytymisen häiriöiden sairausryhmän sekä tuki- ja liikuntaelinten sekäsidekudoksen sairausryhmän. Väärän hälytyksen todennäköisyys $p = 0.1\%$ ja skaalauskerroin $k = 1000$.

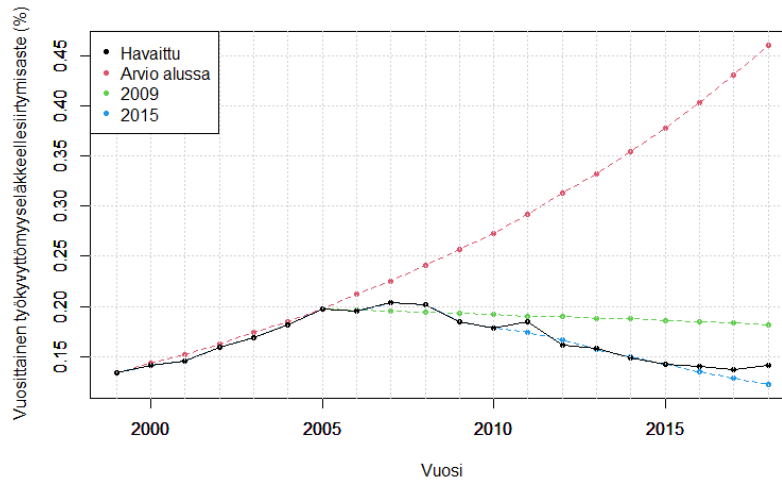
Tarkastellaan ensin näitä ryhmiä kaikilla ikäryhmillä yhteensä. Kuvassa 16 on esitetty trendimuutoksen havaitseminen mielenterveyden ja käyttäytymisen häiriöiden osalta. Vuonna 2010 ja 2015 on havaittu trendimuutokset ja tehty uudet ennusteet, joiden mukaan trendi on muuttunut väheneväksi.



Kuva 15: Trendimuutoksen havaitseminen 18-62 -vuotiaiden työkyvyttömyyseläkkeellesiirtymisasteesta.



Kuva 16: Trendimuutoksen havaitseminen kun sairausryhmänä on mielen terveyden ja käyttäytymisen häiriöt kaiken ikäisillä.



Kuva 17: Trendimuutoksen havaitseminen kun sairauspääryhmänä on tuki- ja liikuntaelinten sekä sidekudoksen sairaudet kaiken ikäisillä.

Kuitenkin lopuksi huomataan havaintojen lähtevän jälleen nousuun, ja menetelmä havaitsee vuonna 2018 jälleen trendimuutoksen.

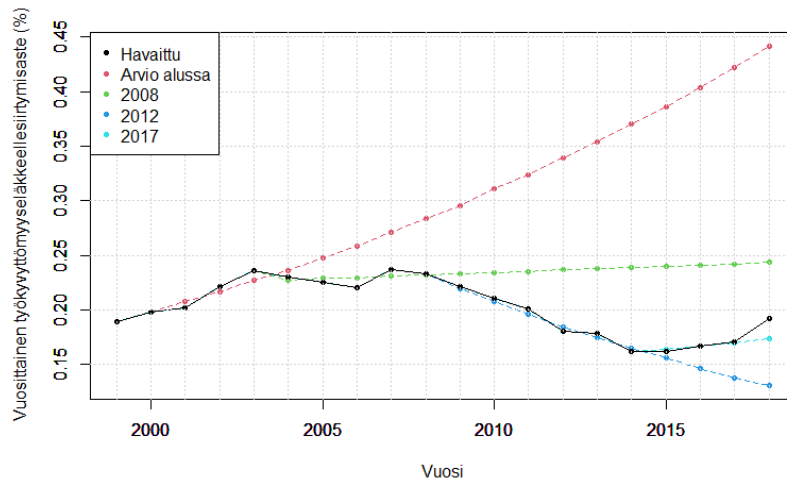
Kuvassa 17 nähdään miten trendimuutos havaitaan kun sairauspääryhmänä on tuki- ja liikuntaelinten sekä sidekudoksen sairaudet. Tässäkin sairauspääryhmässä trendi on ollut ensin nousevaa ja sitten kääntynyt laskuun. Trendimuutokset havaitaan vuosina 2009 ja 2015.

Kuvio toistuu melko samanlaisena molemmissa sairauspäärymissä, ja havainnoissa voidaan nähdä loppua kohden pientä nousua. Mielenterveyden ja käyttäytymisen häiriöiden kohdalla nousu on kuitenkin selkeämpää kuin tuki- ja liikuntaelinten sekä sidekudoksen sairauksien kohdalla.

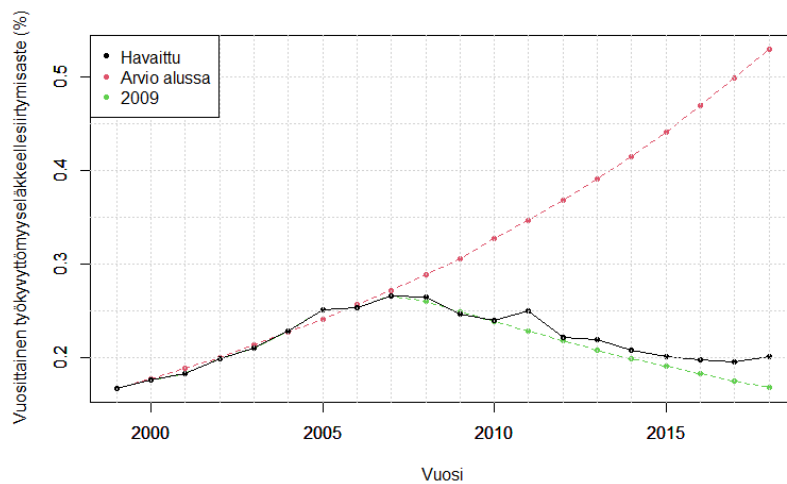
Tarkastellaan vielä näitä kahta sairauspääryhmää 18-62 -vuotiaiden osalta.

Kuvassa 18 nähdään mielenterveyden ja käyttäytymisen häiriöiden sairauspääryhmän trendimuutokset 18-62 -vuotiailla. Kuvio trendimuutokseen on melko samanlainen kuin kuvassa 16. Voidaan kuitenkin huomata että kuvassa 18 %-luvut ovat suurempia.

Kuva 19 näyttää trendimuutoksen havaitsemisen kun sairauspääryhmänä on tuki- ja liikuntaelinten sekä sidekudoksen sairaudet 18-62 -vuotiailla. Tässäkin voidaan huomata %-lukujen kasvaneen. Kuvio on melko samanlainen kuin kuvassa 17, mutta havainnot eivät lähde yhtä suureen laskuun.



Kuva 18: Trendimuutoksen havaitseminen kun sairauspääryhmänä on mielenterveyden ja käyttäytymisen häiriöt 18-62 -vuotiailla.



Kuva 19: Trendimuutoksen havaitseminen kun sairauspääryhmänä on tuki- ja liikuntaelinten sekä sidekudoksen sairaudet 18-62 -vuotiailla.

5 Loppupäätelmät

Menetelmä vaikuttaa havaitsevan hyvin trendin muutoksia kun tarkastellaan työkyvyttömyyseläkkeelle jääneitä, ja kun menetelmän herkkyyttä on säädetty sopivasti skaalauskerroimen sekä väärän hälytyksen todennäköisyyden avulla.

Ikäluokkien koot vaikuttavat työkyvyttömyyseläkkeelle siirtyneiden määrään huomattavasti, erityisesti suurien ikäluokkien kohdalla. Ikävakioidulla aineistolla tämän vaikutuksen voisi poistaa, ja voisi olla mielekästä tutkia miten menetelmä toimii tällä tavalla muokatun datan kanssa. Ikävakiointia varten aineistoa tulisi kuitenkin olla 1-vuotisikäluokittain.

Kuukausikohtaista aineistoa voisi myös olla mielenkiintoista tutkia menetelmän kanssa. Kuukausikohtaisen aineiston kanssa aineistolle tulisi tehdä kausivaihtelun vakiointi, sillä työkyvyttömyyseläkkeiden kohdalla etenkin vuoden alussa sekä syksyn alussa on yleensä havaittavissa huomattavasti enemmän eläkkeelle siirtyneitä kuin muina aikoina.

Shiryaev-Robertsin menetelmää voisi siis käyttää hyödyksi, kun halutaan arvioida tulevien vuosien työkyvyttömyyseläkkeelle siirtyvien lukumääriä. Työssä esitetyistä kuvista voidaan kaikkia ikäluokkia tarkasteltaessa huomata pientä nousua loppua kohden, joka voisi viitata trendin kääntymistä jälleen nousuun. Menetelmä on kuitenkin melko herkkä muutoksille etenkin skaalauskerroimen ja väärän hälytyksen todennäköisyyden ollessa suuria. Tällöin menetelmän tekemiin hälytyksiin tulee suhtautua varauksella, sillä pienetkin satunnaiset muutokset voivat aiheuttaa hälytyksen.

Liite A Ohjelmalistaus

```
rm(list = ls())
#Ladataan tarvittavat paketit
library(sqldf)
library(forecast)
library(parallel)
library(xlsx)
library(readxl)

#SHIRYAEV-ROBERTSIN MENETELMÄ TASON MUUTOKSELLE
#Määritellään ensin funktiot
#Funktio S() saa parametreiksi;
#i = parametrin i arvo Shiryaev-Robertsin havaintoluvussa
#p = kerroin jolla lambda muuttuu
#lambda = alkuperäinen lambdan arvo
#arvot = tarkasteltavat havainnot joista muutos halutaan
#löytää
#Funktio palauttaa Shiryaev-Robertsin havaintoluvun

S<-function(i,p,lambda,arvot){
  a=1;
  summa=0;
  for(k in 1:i){
    eksponentti=0;
    kertoja=1;
    for(j in k:i){
      eksponentti<-eksponentti+(lambda-p*lambda);
      kertoja<-kertoja*((lambda*p)/lambda)^arvot[j];
    }
    summa<-summa+exp(eksponentti)*kertoja;
  }
  return(summa)
}

#Funktio optimiS() saa parametreikseen;
#alaraja = välin alaraja jolla funktiota S() maksimoidaan
#ylaraja = välin yläaraja jolla funktiota S() maksimoidaan
#i = parametrin i arvo Shiryaev-Robertsin havaintoluvussa
#p = tuntematon, jolle S()-funktioita maksimoidaan
#lambda = lambdan arvo ennen muutosta
```

```

#arvot = tarkasteltavat havainnot joista muutos halutaan
#löytää
#Funktio palauttaa arvon p, jolla funktio S() saavuttaa
#suurimman arvonsa määritellyllä välillä
optimiS<-function(alaraja,ylaraja,i,p,lambda,arvot){
  S2<-function(p){
    return(S(i,p,lambda,arvot))
  }
  o<-optimize(S2,interval=c(alaraja,ylaraja),maximum=TRUE);
  palautus<-o$maximum
  if(is.na(o$objective)){
    palautus<-1
  }
  return(palautus)
}

#Tarkasteltavat havainnot ovat vektorissa x
x<-c(16,11,10,12,6,3,2,3,1,3,2,6,3,2,3,2,6,4,4,3)
lambda<-12 #lambda ennen muutosta
tn<-0.01 #väärän hälytyksen todennäköisyys
Kynnysarvo<-vector()
SRtaso<-vector()
lopetus<-0
i=1
#Shiryaev-Robertsin havaintolukuja lasketaan kunnes ne
#ylittävät kynnysarvon
while(lopetus==0){
  if(i>=length(x)){
    lopetus<-1
  }
  numCores<-detectCores()
  cl<-makeCluster(numCores)
  clusterExport(cl=cl, varlist=c("lambda","i","optimiS","S"),
  envir=environment())
  SIMtaso<-unlist(parLapply(cl, 1:10000,function(x){
    arvoja<-rpois(i,12)
    p0<-optimiS(0.01,2,i,p2,lambda,arvoja)
    return(S(i,p0,12,arvoja))
  }), use.names = FALSE)
  stopCluster(cl)
  Kynnysarvo[i]<-min(SIMtaso[SIMtaso>quantile(SIMtaso,1-tn)])
}

```

```

p0<-optimiS(0.01,2,i,p,lambda,x)
SRtaso[i]<-S(i,p0,lambda,x)
if(SRtaso[i]>=Kynnysarvo[i]){
  lopetus<-1
}
i=i+1
}

#SHIRYAEV-ROBERTSIN MENETELMÄ TRENDIMUUTOKSELLE
#Määritellään ensin funktiot
#ShiryaevRobertsTrendi()-funktio saa parametreikseen;
#data = aineisto jonka trendinmuutosta tarkkaillaan
#alpha = alkuperäinen muutos datan havaintojen välillä
#tn = väärän hälytyksen todennäköisyys
#k = skaalauskerroin datalle.
#Funktio palauttaa listan;
#indeksi = kertoo kuinka mones datan havainto
#aiheutti hälytyksen
#rhoo = kerroin uuden alphan laskemista varten
#v = hetki jolloin muutos on tapahtunut
#Uusi ennuste = vektori johon on laskettu ennuste
#uudella alphalla
#SR-arvot = Shiryaev-Robertsin sarjojen viimeisiä
#havaintolukuja välillä [1,indeksi]
#Kynnysarvot = vektori kynnysarvoista
#jokaiselle n=1,...,indeksi
#Pnollat = rhoon arvot jokaiselle n=1,...,indeksi
#V = vektori muuttujan v arvoista jokaisella
#n=1,...,indeksi

ShiryaevRobertsTrendi<-function(data, alpha, tn,k){
  Srarvot<-vector();
  if(length(data)<=1){
    #Jos datan koko on 1 tai pienempi,
    #ei trendimuutosta havaita
    print("Ei hälytystä!")
    lista<-list("indeksi"="tyhjä","rhoo"=0,"v"=1,
    "Uusi ennuste"=0, "Uudet lambdat"=0,
    "SR-arvot"=0, "Kynnysarvot"=0, "Pnollat"=0,
    "V"=0)
    return(lista);
  }

```

```

}
veet<-vector();
ksiit<-vector();
rhoot<-vector();
kynnysarvo<-vector();
lambdat<-c(data[1]);
for(i in 2:length(data)){
  lambdat[i]<-lambdat[i-1]*alpha;
}
for(n in 1:length(data)){
  numCores<-detectCores()
  cl<-makeCluster(numCores)
  clusterEvalQ(cl, {})
  clusterExport(cl, varlist=c("data",
  "n","lambdat", "optimiKsii",
  "ksii","alpha","k"),
  envir=environment())
  APU<-unlist(parLapply(cl,1:n,function(v){
    p0<-optimiKsii(0.00001,2,n,v,n,p1,alpha/k,
    lambdat/k,data/k)
    ksiit<-vector()
    for(i in 1:n){
      ksiit[i]<-ksii(n,v,i,p0,alpha/k,
      lambdat/k,data/k)
    }
    suurinKsii<-max(ksiit)
    return(c(v,p0,suurinKsii))
  })))
  stopCluster(cl)
  Vearvot<-APU[c(TRUE, FALSE,FALSE)]
  Ksiiarvot<-APU[c(FALSE,FALSE,TRUE)]
  Rhooarvot<-APU[c(FALSE,TRUE,FALSE)]
  indeksi<-which(Ksiiarvot==max(Ksiiarvot))
  #Etsii indeksin jonka kohdalla
  #Ksiiarvot-vektorissa on suurin arvo
  veet[n]<-Vearvot[indeksi]
  #muuttujan v arvo jolla ksii saa
  #suurimman arvonsa tarkasteluajanhetkellä n
  ksiit[n]<-Ksiiarvot[indeksi]
  #Suurin ksii tarkasteluajanhetkellä n
  rhoot[n]<-Rhooarvot[indeksi]

```

```

#rhoon arvo jolla ksii saa suurimman
#arvonsa tarkasteluajanhetkellä n
#Kynnysarvon laskeminen; rinnakkaislaskennalla
#simuloidaan 10,000 kertaa tilanne jossa
#trendi ei muutu
numCores<-detectCores()
cl<-makeCluster(numCores)
clusterEvalQ(cl, {})
clusterExport(cl=cl,
varlist=c("data","n","lambdat","optimiKsii",
"ksii","alpha","k"), envir=environment())
SIM<-unlist(parLapply(cl, 1:10000,function(x){
  arvoja<-vector();
  for(i in 1:length(data)){
    arvoja[i]<-rpois(1,lambdat[i])
  }
  ksiit1<-vector();
  for(v in 1:n){
    p0<-optimiKsii(0.01,2,n,v,n,p2,alpha/k,
    lambdat/k,arvoja/k)
    ksiit1[v]<-ksii(n,v,n,p0,alpha/k,
    lambdat/k,arvoja/k)
  }
  return(max(ksiit1))
}), use.names = FALSE)
stopCluster(cl)
kynnysarvo[n]<-min(SIM[SIM>quantile(SIM,1-tn)])
if(kynnysarvo[n]==Inf){
  kynnysarvo[n]<-ksiit[n]+0.1
}
#Valitaan halutun prosenttiosuuden pienin arvo
#simuloiduista arvoista
if(ksiit[n]>=kynnysarvo[n]){
#Tapahtuu hälytys ja tehdään ennuste
pn1=rhoot[n] #rhoon arvo
vn1=veet[n] #ajanhetki
#Luodaan ennuste, ennusteen laskemisen
#ajaksi lambdan arvot on kerrottu arvolla
#1000 jotta ennuste on tasaisempi
Ennuste<-vector()
lambdat2<-lambdat;

```

```

if(vn1==1){
  lambdat2[1]<-lambdat2[1]*pn1;
  Ennuste[1]<-rpois(1,lambdat2[1]*1000)
  for(i in 2:length(data)){
    lambdat2[i]<-lambdat2[i-1]*alpha*pn1;
    Ennuste[i]<-rpois(1,lambdat2[i]*1000)
  }
}else{
  for(i in 1:length(data)){
    if(i<vn1){
      Ennuste[i]<-data[i]*1000
    }else{
      lambdat2[i]<-lambdat2[i-1]*alpha*pn1;
      Ennuste[i]<-rpois(1,lambdat2[i]*1000)
    }
  }
}
Ennuste<-Ennuste/1000
lista<-list("indeksi"=n,"rhoo"=pn1,"v"=vn1,
"Uusi ennuste"=Ennuste,
"Uudet lambdat"=lambdat2,
"SR-arvot"=ksiit,
"Kynnysarvot"=kynnysarvo,
"Rhoot"=rhoot, "V"=veet)
return(lista)
}
if(n==length(data) && ksiit[n]<kynnysarvo[n]){
  print("Ei hälytystä!")
  lista<-list("indeksi"="tyhjä","rhoo"=1,
"v"=n, "Uusi ennuste"=0, "Uudet lambdat"=0,
"SR-arvot"=ksiit, "Kynnysarvot"=kynnysarvo,
"Rhoot"=rhoot, "V"=veet) return(lista)
}
}
}

#optimiKsii()-funktio saa parametreiksi;
#alaraja = välin alaraja jolla ksii()-funktiota
#maksimoidaan
#ylaraja = välin yläaraja jolla ksii()-funktiota
#maksimoidaan

```

```

#n = tarkasteltavan datan koko
#v = ajanhetki jolloin muutos tapahtuu
#i = parametrin i arvo Shiryaev-Robertsin havaintoluvussa
#p = on tuntematon, jolle ksii()-funktiota maksimoidaan
#alpha = alkuperäinen muutos havaintojen välillä
#lambdat = lambdojen arvo kun muutosta ei ole tapahtunut
#arvot = vektori havainnoista joista trendimuutos halutaan
#löytää
#optimiKsii()-funktio palauttaa parametrin p arvon,
#jolla funktio ksii()-saa suurimman arvonsa

optimiKsii<-function(alaraja,ylaraja,n,v,i,p,
alpha,lambdat,arvot){
  ksii2<-function(p){
    return(ksii(n,v,i,p,alpha,lambdat,arvot))
  }
  o<-optimize(ksii2,interval=c(alaraja,ylaraja),maximum=TRUE);
  palautus<-o$maximum
  if(is.na(o$objective)){
    palautus<-1
  }
  return(palautus)
}

#ksii()-funktio saa parametreiksi;
#n = tarkasteltavan datan koko
#v = ajanhetki jolloin muutos tapahtuu
#i = parametrin i arvo Shiryaev-Robertsin havaintoluvussa
#p = intensiteetin muutoskerroin
#alpha = alkuperäinen muutos havaintojen välillä
#lambdat = lambdojen arvo kun muutosta ei ole tapahtunut
#arvot = vektori havainnoista joista trendimuutos halutaan
#löytää
#ksii()-funktio palauttaa Shiryaev-Robertsin sarjan luvun

ksii<-function(n,v,i,p,alpha,lambdat,arvot){
  a=1;
  summa=0;
  lambdapilkut<-lambdat
  if(v<n){
    for(j in v:n){

```



```

    lambdapilkut[j]<-lambdapilkut[j]*(p^a);
    a=a+1;
  }
}
for(k in 1:i){
  eksponentti=0;
  kertoja=1;
  for(j in k:i){
    eksponentti<-eksponentti+(lambdat[j]-lambdapilkut[j]);
    kertoja<-kertoja*((lambdapilkut[j]/lambdat[j]))^arvot[j];
  }
  summa<-summa+exp(eksponentti)*kertoja;
  if(is.na(summa+exp(eksponentti)*kertoja)){
    summa<-summa+0
  }else{
    summa<-summa+exp(eksponentti)*kertoja;
  }
}
return(summa)
}

```

```

#Tuodaan data, ensin väkiluku vuosina 1999-2018
#ja sitten tiedot työkyvyttömyyseläkkeelle
#jääneistä vuosina 1999-2018. vakiluku.xlsx
#sisältää Suomen väkiluvut eri ikäluokittain

```

```

vakilukukaikki<- read_excel("vakiluku.xlsx",
sheet = "YHTEENSA")
elakkeet<-read_excel("elakkeet.xlsx")

```

```

#elakkeet.xlsx sisältää työkyvyttömyyseläkkeelle
#jääneet eri luokissa kuukausittain vuosien
#1999-2018 aikana

```

```

vuositaso<-sqldf('select vuosi, sum(lkm)
as total from elakkeet where ika="9"
and sex="9" and laji="9" and
diag="9" group by vuosi')

```

```

#Tässä tarkastellaan kaikkia työkyvyttömyys-

```

```

#eläkkeelle jääneitä yhteensä
#Lasketaan väestöön suhteutettu työkyvyttömyys-
#eläkkeelle jääneiden lukumäärä ja kerrotaan se
#luvulla 100 ja skaalauskerroimella k

k=1000
vakiluku<-vakilukukaikki$ika9
#Valitaan oikean ikäluokan väkiluvut
suhteutettu<-vuositaso$total/vakiluku
suhteutettuts<-ts(suhteutettu*100*k, start = c(1999),
end = c(2018), frequency = 1) #Luodaan aikasarja

#Luodaan ennuste ajalle ennen varsinaisia havaintoja.
#Ennusteen avulla määritellään alphan arvo aluksi

data<-suhteutettuts
ylaraja<-7 #Valitaan kuinka monta arvoa aineistosta
#otetaan ennusteen tekoa varten
x<-data[1:ylaraja]
tsx<-ts(x,start = c(1999),end = c(1999+ylaraja-1),
frequency = 1)
revx<-ts(rev(tsx),frequency = 1) #Käännetään data
fc<-forecast(auto.arima(revx),ylaraja) #Tehdään ennuste
fc$mean<-ts(rev(fc$mean),end=tsp(tsx)[1]-1, frequency = 1)
#Käännetään data uudestaan ja luodaan aikasarjan
fc$upper<-fc$upper[ylaraja:1,]
fc$lower<-fc$lower[ylaraja:1,]
fc$x<-tsx
uusi<-ts(c(fc$mean,data[1:ylaraja]), start=c(1999-ylaraja),
end = c(1999+ylaraja-1),frequency = 1)
#Yhdistetään ennuste ja data ylärajaan asti
kerroin<-vector()
for(i in 1:length(uusi)-1){
  kerroin[i]<-(uusi[i+1]-uusi[i])/uusi[i]
}
#Lasketaan kuinka paljon arvot kasvavat
#suhteessa edelliseen
b=mean(kerroin)
a=1+b #Alphan arvo aluksi
lambda<-data[1]/a #lambdanolla
#Luodaan alkuennuste:

```

```

alkuennuste<-vector()
lambdat<-vector()
for(i in 1:20){
  alkuennuste[i]<-rpois(1,lambda*(a^(i))*1000)
  lambdat[i]<-lambda*(a^i)
}
#alkuennusteen siloittamiseksi lambdat kerrotaan arvolla
#1000
alkuennuste<-alkuennuste/(k*1000)
alkuennustets<-ts(alkuennuste,start = c(1999), end =
c(2018),frequency = 1)
uusidata<-data
tn<-0.1 #Väärän hälytyksen todennäköisyys desimaalilukuna
tulokset<-list() #Tähän tallennetaan
#Shiryaev-Robertsin menetelmän tulokset
ennusteet<-list() #Tähän tallennetaan
#Shiryaev-Robertsin menetelmän avulla tehdyt
#uudet ennusteet
indeksi<-c(1)
Vuosiluvut<-vector()
minimit<-c(min(alkuennustets),min(suhteutettuts/k))
maksimit<-c(max(alkuennustets),max(suhteutettuts/k))
lopetus<-0
i=1
#Toistetaan Shiryaev-Robertsin menetelmää
#kunnes ollaan käyty läpi koko aineisto .
#Tulokset ja ennusteet tallennetaan listoiksi.
while(lopetus==0){
  tulos<-ShiryaevRobertsTrendi(uusidata,a,tn,k)
  if(tulos$indeksi=="tyhjä"){
    tulokset[[i]]<-tulos lopetus<-1
  }else{
    tulokset[[i]]<-tulos
    indeksi[i+1]<-tulos$indeksi
    ennuste<-ts(c(data[1:(indeksi[i]+tulos$v-2)],
tulos$'Uusi ennuste'[tulos$v:
length(tulos$'Uusi ennuste')]), start=c(1999),
end = c(2018),frequency = 1)
    minimi[i+2]<-min(ennuste/k)
    maksimi[i+2]<-max(ennuste/k)
    indeksi[i+1]<-indeksi[i]+tulos$indeksi-1
  }
}

```

```

    if(indeksi[i+1]>20){
      lopetus<-1
    }
    ennusteet[[i]]<-ennuste/k
    a<-a*tulos$rhoo
    uusidata<-uusidata[(tulos$indeksi) :length(uusidata)]
    i=i+1
  }
}

#Tehdään kuva:
rajat<-c(min(minimit),max(maksimit))
plot(alkuennustets,type="o",col=2,pch=20,
lty=2,ylim=rajat,xlim=c(1999,2018),ylab="",xlab="")
par(new=T)
for(i in 1:(length(indeksi)-1)){
  plot(ennusteet[[i]],type="o",col=i+2,pch=20,lty=2,
ylim=rajat,xlim=c(1999,2018),ylab="",xlab="")
  par(new=T)
  Vuosiluvut[i]<-vuositaso$vuosi[indeksi[i+1]]
}
plot(data/k,type="o",col=1,pch=20,lty=1,
ylim=rajat,xlim=c(1999,2018),ylab="Vuositainen
työkyvyttömyyseläkkeellesiirtymisaste (%)",xlab="Vuosi")
par(new=T)
abline(v=c(1999:2018),col="lightgray",lty="dotted")
abline(h=seq(floor(rajat[1]),ceiling(rajat[2]),0.1),
col="lightgray",lty="dotted") legend("topleft",legend =
c("Havaittu","Arvio alussa",Vuosiluvut),
col=c(1,2,c(2+1(length(indeksi)-1))),
pch=rep(20,length(indeksi)+1),text.col = "black")

```

Kirjallisuutta

- [1] D Abgrall, M Habart, C Rainer, A Sow (2018) Exploring the longevity risk using statistical tools derived from the Shiryayev-Roberts procedure. *European Actuarial Journal* 8:27-51. <https://doi.org/10.1007/s13385-018-0168-4>
- [2] AS Polunchenko, AG Tartakovsky (2010) On optimality of the Shiryayev-Roberts procedure for detecting a change in distribution. *The Annals of Statistics* 38(6):3445-3457. <https://www.jstor.org/stable/29765270>
- [3] J Rantala, M Hietaniemi, H Nyman, M Laaksonen, S Kuivalainen (2017) Työkyvyttömyyseläkkeensaajien eläketurva ja toimeentulo 2000-luvulla. *Eläketurvakeskuksen tutkimuksia*. https://www.etk.fi/wp-content/uploads/Tyokyvyyttomyyselakkeensaajien_elaketurva_ja_toimeentulo_2000_luvulla.pdf
- [4] M Pollak, AG Tartakovsky (2009) On optimality properties of the Shiryayev-Roberts procedure. *Statistica Sinica* 19(4): 1729-1739. <https://www.jstor.org/stable/24308927>
- [5] SW Roberts (1966) A Comparison of Some Control Chart Procedures. *Technometrics* 8(3):411-430. <https://www.jstor.org/stable/1266688>
- [6] AN Shiryayev (1963) On optimum methods in quickest detection problems. *Theory of probability and its applications* 8(1):22-46
- [7] Suomen virallinen tilasto (SVT): Väestörakenne [verkkojulkaisu]. ISSN=1797-5379. Helsinki: Tilastokeskus [viitattu: 3.4.2020]. Saantipa: <http://www.stat.fi/til/vaerak/tau.html>
- [8] M Laaksonen (2019) Työkyvyttömyyseläkkeelle siirtyneiden määrä kääntyi kasvuun 2018. *Eläketurvakeskuksen blogit* <https://www.etk.fi/blogit/tyokyvyyttomyyselakkeelle-siirtyneiden-maara-kaantyi-kasvuun-2018/>