

Saara Vehkala

**Syklinen synkronia – sovellus metsämyyrien
lisääntymiseen**

Tilastotiede
pro gradu -tutkielma
30. marraskuuta 2007

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Jyväskylä

Tekijä: Saara Vehkala

Yhteystiedot: saara.vehkala@jyu.fi

Työn nimi: Syklinen synkronia – sovellus metsämyyrien lisääntymiseen

Title in English: Syclic synchrony – an application to reproduction of bank voles

Työ: Tilastotiede pro gradu -tutkielma

Sivumäärä: 59

Tiivistelmä: Tutkielman teoriaosassa on laskettu kahden, kolmen ja neljän havainnon välisten erotusten pistetodennäköisyysjakaumat riippumattomassa tilanteessa syklisessä ajassa. Empiirisessä osassa on verrattu metsämyyräotosten vastaavia erotuksia tähän nollahypoteesijakaumaan ja tutkittu, esiintyykö aineistossa synkroniaa ilmentävää riippuvuutta. Lisäksi synkronian jakautumista otoksen sisällä on tutkittu multinomijakauman avulla.

English abstract: In this study there have been calculated distributions for differences between two, three or four objects in cyclic time. A field data of bank voles has been compared with them to analyze if there's synchrony in it. Additionally, the division of synchrony inside the samples has been analyzed by using multinomial distribution.

Avainsanat: synkronia, metsämyyrä

Keywords: synchrony, bank vole

Sisältö

1	Esipuhe	1
2	Johdanto	3
3	Syklinen aika	5
3.1	Kaksi havaintoa syklisessä ajassa	5
4	Kolmen havainnon välisten erotusten teoreettinen jakauma	7
4.1	Erotusten d_1 ja d_2 mahdolliset arvot	7
4.2	Erotusten pistetodennäköisyysjakauman määrittäminen	8
4.3	Pienin erotus d_1	9
4.3.1	d_1 :n arvojen frekvenssien laskeminen	9
4.3.2	d_1 :n frekvenssijakauma	12
4.3.3	Kaikkien havaintokombinaatioiden kokonaismäärä	15
4.4	Toiseksi pienimmän erotuksen d_2 jakauma	17
4.4.1	Nollan ja ykkösen mittaisten erotusten frekvenssit	17
4.4.2	d_2 :n arvojen frekvenssien laskeminen ja todennäköisyysjakauma	17
5	Simuloinnin tuottama erotusten jakauma kolmen havainnon otoksille	29
5.1	Simulointi kolmen havainnon otoksille ja tulosten vertailu teoriaan	29
5.2	Simulointimallin luotettavuus	30
6	Neljän havainnon välisten erotusten jakauma	34
6.1	Erotusten d_1 , d_2 ja d_3 jakauma simuloimalla	34
7	Empiirinen tutkimus	36
7.1	Aineisto	36
7.2	Synkronian määrittely	36
7.3	Synkronian mittaaminen	37
7.4	Satunnaiset erotukset kolmen ja neljän havainnon välillä	38
7.5	Kahden havainnon välinen synkronia	39
7.6	Kolmen havainnon välinen synkronia	40
7.7	Neljän havainnon välinen synkronia	40

8	Synkronian tutkiminen aineistoa luokittelemalla	45
8.1	Teoria kolmen havainnon otosten luokittelulle	45
8.2	Synkronialuokkien todennäköisyyksien laskeminen	46
8.3	Luokittelumenetelmän soveltaminen aineistoon	48
9	Johtopäätökset	49
Liitteet		
A	Aineisto	51
B	Kolmen havainnon otosten simulointiin käytetty R-koodi	53
	Viitteet	55
	Kuvat	56

1 Esipuhe

Sain metsämyyräaiheeni tilastotieteen kandidaattityötä varten kevättalvella 2006. Harri Högmander ja Kari Nissinen olivat järjestäneet mahdollisuuden saada tutkittavaksi oikeaa dataa. Aiheenjakoon oli kutsuttu biologian ja ympäristötieteen laitokselta Tapio Mappes ja Esa Koskela. Meitä aiheen tarvitsijoita oli enemmän kuin mihin he olivat varautuneet, joten he rupesivat varovasti muistelemaan: ”Olisihan meillä vielä yksi mahdollisuus, sellainen vähän luovuutta vaativa, erikoisempi juttu...” Kiinnostuin ja tein siitä LuK-tutkielman mutta aiheen käsittely tuntui jäävän aivan kesken. Niinpä Harri Högmander pyynnöstäni tarkisti, olisiko siitä graduaiheeksi, ja olihan siitä.

Olen esitellyt aihetta tilastotieteen LuK- ja graduseminaarissa sekä matematiikan graduseminaarissa. Aihe on kiinnostanut monia kuulijoita, kuten toista työn tarkastajaa Jukka Nyblomia, ja olemme ilman mitään biologista koulutusta keksineet hurjia teorioita siitä, miksi data näyttää siltä, miltä se näyttää. Valitettavasti näitä ideoita ei voi tieteelliseen julkaisuun sisällyttää, joten tutkailkaa itse ja innostukaa! Suositelen histogrammeja 7.1, 7.3 ja 7.5, joista näkyy kaikki tarpeellinen. Teoriaosaan syventymistä en suosittele kenellekään elämänhaluiselle; kysykää sen tarkistaneelta Harri Högmanderilta, miksi. Kuten jokainen tutkija, toivon toki kehittämäni menetelmien ja tuloksien koituvan muidenkin hyödyksi. Kimppuun siis, sekä myyrien että muiden ehkä-synkronisten elikoiden tutkijat!

Saara Vehkala



Kuva 1.1: Metsämyyrä Matti Laineen kuvaamana

2 Johdanto

Tutkielman teoriaosan tavoitteena on kehittää menetelmä, jolla voidaan tutkia yksittäisten, ajan suhteen mitattujen havaintojen riippuvuutta toisistaan syklisessä ajassa. Riippuvuusilmiön todentaminen perustuu vertailuun riippumattoman tilanteen suhteen, jossa havainnot jakautuvat täysin satunnaisesti.

Synkronian tutkiminen perustuu pisteiden välisten erotusten tarkkailuun: onko synkroniaa todentavia erotuksia enemmän kuin satunnaisessa tilanteessa? Tätä varten on laskettu riippumattoman tilanteen mukainen pistetodennäköisyysjakauma kahden, kolmen tai neljän havainnon välillä oleville yhdelle, kahdelle tai kolmelle pienimmälle erotukselle. Kahden ja kolmen havainnon välillä olevien erotusten jakaumat on laskettu teoreettisesti mutta neljän havainnon välillä olevien kolmen pienimmän erotuksen jakaumaa on vain simuloitu.

Menetelmiä on sovellettu todelliseen havaintoaineistoon. Lisäksi kolmen ja neljän havainnon otoksista koostuvaa aineistoa on tutkittu satunnaisesti valittujen erotusten perusteella, sillä nämä ovat riippumattomia toisin kuin pienimmät mahdolliset otoksen sisäiset erotukset. Myös synkronian jakautumista otosten sisällä on tutkittu kolmen havainnon otoksille.

Empiirisen tutkimuksen aiheena on eläinten lisääntymisen syklinen vaihtelu ja kohteena ovat metsämyyrän luonnonpopulaatiot Keski-Suomessa. Tavoitteena on selvittää, synnyttävätkö lähekkäin olevat metsämyyrät synkroniassa tai ovatko niiden kiimasyklit synkroniassa. Kolmas biologisesti perusteltavissa oleva vaihtoehto olisi epäsynkronia eli säännöllisin väliajoin tapahtuva lisääntyminen. Sitä ei kuitenkaan tutkita tässä opinnäytetyössä, sillä sitä ilmentävien erotusten tarkka määrittäminen biologisin perustein on vaikeaa ja voi sekoittua kiimasykliin synkroniaan. Synkroniaa tutkitaan yksilökohtaisena, päivän tarkkuudella esiintyvänä ilmiönä, jonka ei kuitenkaan tulisi olla seurausta samanlaisista luonnonoloista, esim. lisääntymiskauden yhtäaikaisesta alkamisesta keväällä.

Tällaista lisääntymisen synkronoitumista on tutkittu vähän eivätkä tulokset ole yhteneviä. Rolf Anker Ims on todennut vuonna 1988 julkaistussa artikkelissa [1], että ei-luonnonoloista johtuvaa lisääntymisen synkroniaa on todettu laboratorio-oloissa muutamalle jyrsijälajille. Kenttäkokein asiaa ei ole tutkittu tähänkään mennessä ennen Tanja Poikosen ym. vuonna 2007 julkaisemaa tutkimusta, jossa löydettiin tukea synkronian esiintymiselle [5]. Artikkelissa esitetään biologisina perusteina synkronialle esim.

suojaa infantisoivia lähialueen naaraita vastaan. Metsämyyränaaraat voivat nimittäin tappaa toistensa poikasia, elleivät ne itse ole imettävinä. Epäsynkronian etuna taas olisi esimerkiksi ravintoresurssien parempi riittävyys, kun kukin naaras poikueineen olisi ravinnontarpeensa maksimivaiheessa eri aikaan.

Ims esittää artikkelissaan [1] oletuksen siitä, että heterogeenisessä ympäristössä epäsynkroniaa voi esiintyä tiheissä populaatioissa sen takia, että ravinnon laadulliset erot reviirien välillä kasvavat. McShean vuonna 1989 julkaistu artikkeli [4] antaa aiheutta olettaa, että tällöin tutkitussa myyräaineistossa (*Microtus pennsylvanicus*) tiheissä populaatioissa esiintynyt epäsynkronia ei johtuisi pelkästään tästä tai myöskään kevään alussa mahdollisesti yhtäaikaisesta lisääntymiskauden aloittamisesta. Tulos ei ole tarpeeksi vahva, mutta toisaalta aineiston sanotaan olevan selvästi sellainen, jossa synkroniaa ei ilmene. Johannesenin vuonna 2000 julkaisemassa artikkelissa [2] on löydetty enemmän synkroniaa harvassa olevilla myyränaarailta (*Microtus oeconomus*) kuin tiheässä olevilla ja lisäksi lähekkäin olevien naaraiden keskuudessa esiintyi epäsynkroniaa, kuten McSheankin tutkimassa aineistossa. Lambinin vuonna 1993 julkaisemassa tutkimuksessa [3] taas ei löydetty yhteyttä populaatiotiheyden ja synkronian välillä. Tosin tällöin on saatu tilastollisesti merkitsevä tulos, jonka mukaan lähekkäin elävät ja läheisessä sukulaissuhteessa olevat naaraat (*Microtus townsendii*) ovat huomattavasti vahvemmin synkronisoituneita kuin muut naaraat pareittain tarkasteltuna. Yleisesti tarkasteltuna tässä tutkimuksessa ei havaittu synkroniaa. Näiden kolmen artikkelin perusteella voisi siis pitää oletettavana mutta ei todistettuna sitä, että epäsynkroniaa esiintyy vahvemmin, kun myyräpopulaatio on tiheä.

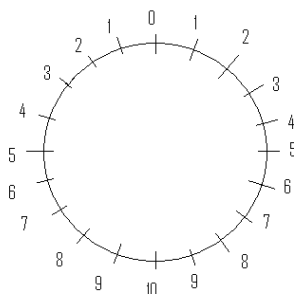
Lisääntymisen synkroniaa käsittelevien tutkimusten vertailu on hankalaa, sillä tutkimuksissa on käytetty hyvin erilaisia metodeja ja oletuksia synkronian ja epäsynkronian määrittämiseen. Useat synkronian määrittämisen menetelmät vaikuttavat puutteellisilta tilastollisen pätevyyden kannalta. Lisäksi koeolosuhteet ja menetelmät ovat hyvin erilaisia ja myös tarkasteltavat myyrälajit vaihtelevat. Missään tämän työn lähdeluetteloon sisältyvistä artikkeleista ei myöskään ole mainittu kiimasykliin synkroniaa, jota taas tässä tutkimuksessa käsitellään rinnan synnytysten synkronian kanssa.

3 Syklinen aika

Määritellään syklinen aika, jossa ajanlasku aloitetaan alusta T yksikön välein. Tällöin aikaa voidaan havainnollistaa lukusuoran sijasta ympyrällä. Syklin pituus on T , jonka kohdalta kierros alkaa alusta: $T = 0$, $T + 1 = 1$ jne. Suurin aikapisteiden välillä oleva välimatka on näin ollen $\lfloor \frac{T}{2} \rfloor$ (luvun kokonaislukuarvo), sillä erotus mitataan lyhintä mahdollista reittiä ympyrän kehää pitkin. Havainnot, kuten niiden erotuksetkin, ilmoitetaan kokonaislukuina. Syklin avulla mitataan pisteen etäisyyttä nolasta, joten $\frac{T}{2}$:ta suuremmat luvut peilataan $\frac{T}{2}$:n suhteen siten, että T :n ollessa parillinen $\frac{T}{2} + 1 = \frac{T}{2} - 1$, $\frac{T}{2} + 2 = \frac{T}{2} - 2, \dots, T - 1 = 1$, $T = 0$ ja T :n ollessa pariton syklin puolivälissä on kahdesti arvo $\lfloor \frac{T}{2} \rfloor$ ja siitä eteenpäin luvut pienenevät nolaa kohti vastaavasti kuin edellä. Kuvassa 3.1 on sykli, jonka maksimi on $T = 20$. Syklin puolivälin eli kymmenen kohdalla numerot alkavat pienentyä, sillä esim. numero 11 on vain 9 yksikön päässä nolasta, koska kiertosuunnalla ei ole merkitystä. Koska T on parillinen, esiintyy maksimiarvo 10 vain kerran. Parittoman T :n tapauksessa maksimiarvo esiintyisi kahdesti. Pidetään järkevänä alarajana T :lle nelosta; $T \geq 4$.

3.1 Kaksi havaintoa syklisessä ajassa

Määritetään kahden syklisessä ajassa olevan aikahavainnon välisen erotuksen todennäköisyysjakauma. Jos havainnot ovat toisistaan riippumattomia eli kaikilla arvoilla on sama esiintymistodennäköisyys, kaikkia erotuksia väliltä $[0, \lfloor \frac{T}{2} \rfloor]$ esiintyy lineaarisessa ajassa keskimäärin yhtä paljon, $\frac{1}{T}$. Syklisessä ajassa taas nolaa esiintyy puolet vähemmän kuin muita yksittäisiä erotuksia. Lisäksi, jos T on parillinen, myös $\frac{T}{2}$:n todennäköi-



Kuva 3.1: Sykli, jossa $T = 20$

syys on vain puolet muista yksittäisistä todennäköisyyksistä. Tämä ilmenee siitä, että syklissä kaikki erotukset esiintyvät kahdesti paitsi nolla ja $\frac{T}{2}$ vain kerran. Mikäli yksi havainto on nollassa ja toinen havainto voi sijoittua yhtä suurella todennäköisyydellä mihin tahansa syklin pisteistä, pisteen lukuarvo ilmoittaa havainnon erotuksen nollassa sijaitsevaan havaintoon. Näin siis kunkin erotuksen todennäköisyys on suoraan verrannollinen vastaavan numeron lukumäärään syklissä. Pistetodennäköisyyksien summa on oltava yksi, mistä johtuen kahden havainnon välisen erotuksen pistetodennäköisyysjakauma riippumattomassa tilanteessa on

$$p(d) = \begin{cases} \frac{1}{T} & \text{kun } d = 0 \text{ tai } d = \frac{T}{2} \\ \frac{2}{T} & \text{kun } d \in [1, \lfloor \frac{T-1}{2} \rfloor]. \end{cases}$$

4 Kolmen havainnon välisten erotusten teoreettinen jakauma

Tutkitaan kolmen syklistä ajassa sijaitsevan havainnon välisiä välimatkoja eli erotuksia. Kolmen havainnon välillä on kolme pareittaista erotusta. Nämä eivät kuitenkaan ole toisistaan riippumattomia kuten itse havainnot ovat, sillä yksi erotus määräytyy täysin kahdesta muusta ollen niiden summa (jos summa on yli $\frac{T}{2}$, se lisäksi peilataan syklin puolivälin suhteen). Tarkastellaan siksi vain osaa erotuksista. Satunnaisesti valitut kaksi erotusta joukosta $x_2 - x_1$, $x_3 - x_1$ ja $x_3 - x_2$ ovat toisistaan riippumattomia, ja niitä voidaan tutkia vastaavasti kuten kahden havainnon välisiä erotuksia. Tätä menetelmää käytetään luvussa 7.4. Voitaisiin myös sopia ensimmäinen havainto vertailupisteeksi, johon kahta seuraavaa havaintoa verrataan. Tällöin pitäisi ottaa huomioon pienimmän havainnon noudattama jakauma. Satunnaisesti valittu havainto kolmen joukosta noudattaa tasajakaumaa mutta pienin niistä ei. Valitaan nyt kuitenkin tutkittaviksi pienin ja toiseksi pienin erotus ja merkitään niitä d_1 :llä ja d_2 :lla. Nämä eivät ole toisistaan riippumattomat, koska niiden välillä on suuruusjärjestys. Joissain tilanteissa kuitenkin pienimmät erotukset voivat olla informatiivisimpia tutkitun ilmiön esiintymisen suhteen.

4.1 Erotusten d_1 ja d_2 mahdolliset arvot

Määritetään kolmen havainnon välillä olevien pienimmän ja toiseksi pienimmän erotuksen mahdolliset arvot. Havaintojen välinen pareittainen erotus on aina ei-negatiivinen, sillä havaintojen järjestyksellä ei ole merkitystä. Erotuksista otetaan siis aina itseisarvo. Pienin erotus d_1 on nolla, jos ja vain jos mitkä tahansa kaksi havaintoa ovat samat. Koska d_2 on toiseksi suurin kolmesta pareittaisesta erotuksesta, se voi olla nolla vain, jos myös kolmas havainto on samassa pisteessä kuin kaksi ensimmäistä. Siten

$$\min(d_1) = \min(d_2) = 0.$$

Jos T ei ole jaollinen kolmella, suurin mahdollinen arvo d_1 :lle on $\lfloor \frac{T}{3} \rfloor$, jolloin d_2 on $\lfloor \frac{T}{3} \rfloor$ tai $\lfloor \frac{T}{3} \rfloor + 1$. Jos T on jaollinen kolmella, d_1 voi olla $\frac{T}{3}$, jolloin d_2 on myös $\frac{T}{3}$. (Ks.

kuvat 4.1 ja 4.2.) Siten

$$\max(d_1) = \begin{cases} \lceil \frac{T}{3} \rceil & \text{kun } T \text{ ei jaollinen } 3\text{:lla} \\ \frac{T}{3} & \text{kun } T \text{ jaollinen } 3\text{:lla.} \end{cases}$$

Toiseksi pienin erotus on suurimmillaan, kun kaksi havaintoa ovat mahdollisimman kaukana toisistaan ja kolmas havainto on samassa pisteessä jommankumman aikaisemman havainnon kanssa. Tällöin $d_1 = 0$ ja $d_2 = \lceil \frac{T}{2} \rceil$ ja

$$\max(d_2) = \begin{cases} \lceil \frac{T}{2} \rceil & \text{kun } T \text{ pariton} \\ \frac{T}{2} & \text{kun } T \text{ parillinen.} \end{cases}$$

4.2 Erotusten pistetodennäköisyysjakauman määrittäminen

Määritetään erotusten d_1 ja d_2 jakauma, kun kolme havaintoa ovat toisistaan riippumattomat eli kukin havainto välillä $[0, T - 1]$ on yhtä todennäköinen. Oletetaan, että ensimmäinen havainto on nollassa, merkitään $x_1 = 0$. Oletus selkeyttää tarkasteluja, mutta ei vaikuta tuloksiin.

Selvitetään yksittäisen erotuksen arvon todennäköisyys kaikkien erotusten d_1 tai d_2 arvojen joukossa. Asetetaan haluttu erotuksen arvo k :ksi, annetaan x_2 :n ja x_3 :n saada kaikki mahdolliset arvokombinaatiot syklin kehällä ja lasketaan niiden kombinaatioiden lukumäärä, joissa erotus k esiintyi. Tehdään tämä niin, että asetetaan ensin $x_2 = 0$ ja annetaan x_3 :n saada vuorotellen kaikki arvot välillä $[0, T - 1]$ ja lasketaan samalla, kuinka monessa kohdassa esiintyi erotuksen d_1 tai d_2 arvo k . Siirretään sitten x_2 :ta pykälä eteenpäin, kierrätetään x_3 :ta syklin ympäri ja lasketaan taas tutkittavan erotuksen esiintymisien määrä. Edetään samaan tapaan niin, että x_2 kiertää koko syklin pykälä kerrallaan ja kunkin arvon kohdalla x_3 kiertää syklin. Näin saadaan laskettua valitun erotuksen k frekvenssi kaikkien mahdollisten havaintokombinaatioiden joukossa.

Jotta ei tehtäisi turhaa työtä, käytetään vielä hyväksi tieto, että tilanne on x_2 :n suhteen identtinen yhtä kaukana x_1 :stä ($x_1 = 0$) oleville arvoille molemmilla puolilla sykliä eli $\text{fr}(k \mid x_2 = 1) = \text{fr}(k \mid x_2 = T - 1)$ jne. (Merkintä $\text{fr}(k \mid x_2)$ tarkoittaa erotuksen k frekvenssiä tilanteissa, jossa $x_1 = 0$, x_2 on kiinnitetty ja x_3 saa vuorotellen kaikki arvot syklin kehällä.) Siksi x_2 :ta tarvitseekin kuljettaa vain puoli kierrosta: $x_2 = 0, \dots, \lceil \frac{T}{2} \rceil$. Loppupuolella sykliä olevat arvot ($x_2 = \lceil \frac{T}{2} \rceil + 1, \dots, T - 1$) otetaan huomioon lopuksi, kun lasketaan yhteen kaikki eri x_2 :n arvoilla saadut frekvenssit. Tällöin mukaan lasketaan kahdesti kaikkien niiden tilanteiden frekvenssit, joilla on peilikuvatapaus syklin toisella puolella: $1, \dots, \lceil \frac{T}{2} \rceil$ mutta ei $\frac{T}{2}$. x_3 :a on kuitenkin kuljetettava koko syklin ympäri, sillä kun kaksi havaintoa on jo asetettu kehälle, ei ole sama onko x_3 vasemmalla vai oikealla puolella sykliä, vaikka etäisyys nollasta olisikin sama. Esim.

tilanne $x_1 = 0$, $x_2 = 1$ ja $x_3 = 1$ tai $x_3 = T - 1$: pienin erotus on joko nolla tai yksi. Erotuksen k ($k \in [0, [\frac{T}{3}]]$ tai $k \in [0, [\frac{T}{2}]]$) frekvenssi on siis

$$\begin{aligned} \text{fr}(k) &= \sum_{a=0}^{T-1} \text{fr}(k \mid x_2 = a) \\ &= \begin{cases} \text{fr}(k \mid x_2 = 0) + 2 \sum_{a=1}^{[T/2]} \text{fr}(k \mid x_2 = a) & \text{kun } T \text{ pariton} \\ \text{fr}(k \mid x_2 = 0) + 2 \sum_{a=1}^{T/2-1} \text{fr}(k \mid x_2 = a) + \text{fr}(k \mid x_2 = \frac{T}{2}) & \text{kun } T \text{ parillinen,} \end{cases} \end{aligned}$$

kun on määritelty $\sum_{b=0}^{T-1} \text{fr}(k \mid x_2 = a, x_3 = b) = \text{fr}(k \mid x_2 = a)$.

4.3 Pienin erotus d_1

Määritetään frekvenssit kaikille d_1 :n arvoille välillä $[0, [\frac{T}{3}]]$, kun kolme havaintoa saavat kaikki mahdolliset yhdistelmät syklin kehällä. Menetelmä on selitetty kappaleessa 4.2.

Otetaan selville ensin erikseen nollan esiintymistiheys. Jos x_1 :n lisäksi myös $x_2 = 0$, on d_1 nolla riippumatta x_3 :n arvosta. Siispä d_1 saa arvon nolla T kertaa, kun x_2 pysyy nollassa ja x_3 saa vuorotellen kaikki arvot nollassa $T - 1$:een. Näin ollen $\text{fr}(d_1 = 0 \mid x_2 = 0) = T$. Kun taas x_2 pysyy kiinteänä missä tahansa muussa pisteessä kuin nolla, d_1 on nolla kaksi kertaa x_3 :n kiertäessä koko syklin. Tämä tapahtuu, kun $x_3 = x_2$ tai $x_3 = x_1 (= 0)$. Tällaisia kierroksia on kaikkiaan $T - 1$ kappaletta: $x_2 = 1, \dots, T - 1$. Siispä $\text{fr}(d_1 = 0 \mid x_2 \neq 0) = (T - 1) \cdot 2$. Kun lasketaan yhteen kaikki tilanteet, joissa $d_1 = 0$, saadaan nollaerotusten frekvenssiksi

$$\begin{aligned} \text{fr}_{d_1}(0) &= \text{fr}(d_1 = 0 \mid x_2 = 0) + \sum_{a=1}^{T-1} \text{fr}(d_1 = 0 \mid x_2 = a) \\ &= T + (T - 1) \cdot 2 = 3T - 2. \end{aligned} \tag{4.1}$$

4.3.1 d_1 :n arvojen frekvenssien laskeminen

Määritetään kaikkien erotusten $d_1 = k \in [1, [\frac{T}{3}]]$ esiintymisten lukumäärä eri tilanteissa x_2 :n suhteen, kun x_3 saa vuorotellen kaikki arvot syklin kehällä. Pidetään k :ta periaatteessa yleisenä, mutta silloin, kun eri suuruiset k :t saavat eri frekvenssit, ilmoitetaan frekvenssi k :sta riippuvana. Näin ollen tilanteet on jaettu yleensä kahteen tai kolmeen osaan k :n suhteen.

Kun frekvenssi riippuu k :sta, ensin käsitellään aina rajatapaus, jota suurempaa erotusta valitulla x_2 :n arvolla/arvojoukolla ei esiinny ja sen jälkeen perustilanne eli millaisissa x :ien yhdistelmissä rajatapauستا pienempi mikä tahansa erotus k esiintyy.

Kaikkiaan käsitellään joka kohdassa arvot alueelta $1 \leq k \leq \lfloor \frac{T}{3} \rfloor$. Nollan frekvenssi käsiteltiin esimerkin vuoksi jo luvun 4.3 alussa, joten siihen ei puututa tässä.

$x_2 < k$: Minimierotus k ei voi esiintyä jos $x_2 < k$, sillä tällöin $x_2 - x_1 < k$ eli näiden välillä on siten pienempi erotus kuin k . Koska d_1 on määritelty olemaan pienin erotus havaintojen välillä, d_1 ei voi olla k .

$$\text{fr}_{d_1}(k) = 0 \quad \forall k \in [1, \frac{T}{3}]$$

$x_2 = k$: Rajatapaus: Jos k on tasan $\frac{T}{3}$, on pienin erotus k :n mittainen vain, kun $x_2 = k = \frac{T}{3}$ ja $x_3 = T - k = T - \frac{T}{3}$.

$$\text{fr}_{d_1}(k) = 1, \text{ kun } k = \frac{T}{3}$$

Tutkitaan sitten tilanne, missä $k < \frac{T}{3}$. Tilannetta havainnollistaa kuva 4.3. Kun $x_2 = k$, pienin erotus on k , kunhan x_3 on tarpeeksi kaukana eli vähintään k :n päässä sekä x_1 :stä että x_2 :sta. Erotus on siis k kaikilla muilla x_3 :n arvoilla paitsi alueella $[x_1 - k + 1, x_2 + k - 1] = [1 - k, k + k - 1] = [1 - k, 2k - 1]$ ja tämän välin pituus on $3k - 1$. Näin ollen erotuksia k esiintyy $T - (3k - 1) = T - 3k + 1$ kertaa.

$$\text{fr}_{d_1}(k) = T - 3k + 1 \text{ kun } k < \frac{T}{3}$$

$x_2 = k + 1, \dots, T - 2k - 1$: Kuva 4.4. Kun $x_2 < T - 2k$, mahtuu x_2 :n jälkeen ennen T :tä kahteen paikkaan x_3 siten, että pienempi erotuksista $x_1 - x_3$ ja $x_2 - x_3$ on arvoltaan k : $x_3 = x_2 + k$ ja $x_3 = x_1 - k$. Koska $k < x_2 - x_1 < 2k$, k :n mittainen minimierotus ei voi esiintyä heti x_1 :stä eteenpäin olevalla välillä. Jotta käsiteltävän alueen yläraja $T - 2k - 1$ olisi vähintään yhtä suuri kuin $k + 1$ eli alueen alaraja, on toteutettava yhtälö $T - 2k - 1 > k + 1 \Leftrightarrow k < \frac{T}{3} - \frac{2}{3}$.

$$\text{fr}_{d_1}(k) = 2 \text{ kun } k < \frac{T}{3} - \frac{2}{3}$$

$x_2 = T - 2k$: Kuva 4.5. Kun $x_2 = T - 2k$ (siis x_2 on $2k$ askelta T :stä vastapäivään), x_2 :n jälkeen tilaa ennen T :tä on $2k$:n verran. x_3 voi olla siis vain yhdessä paikassa, jotta minimierotus olisi k : $x_3 = T - k = x_2 + k$. (Tällöin k :n mittaisia erotuksia on jopa kaksi.) Minimierotus on tällöin k , sillä $k < x_2 - x_1 < 2k$. Jos k on alle $\frac{T}{4}$, kohdassa $T - 2k$ x_2 menee yli syklin puolivälin, ja jos k on $\frac{T}{3}$, kohdassa $T - 2k$ x_2 on k , joten näitä tilanteita ei käsitellä tässä.

$$\text{fr}_{d_1}(k) = 1 \text{ kun } \frac{T}{4} < k < \frac{T}{3}$$

$x_2 = T - 2k + 1, \dots, 2k - 1$: Kuva 4.6. Jos $x_2 > T - 2k$, ei minimierotusta k mahdu olemaan missään.

$$\text{fr}_{d_1}(k) = 0 \quad \forall k \in [1, \frac{T}{3}]$$

$x_2 = 2k$: Kun $k = \frac{T}{4}$ ja x_2 on syklin puolivälissä eli $2k$:ssa, molemmille puolille jää $2k$:n pituiset pätkät, joten minimierotus k esiintyy, kun $x_3 = k$ tai $x_3 = 3k$.

$$\text{fr}_{d_1}(k) = 2 \quad \text{kun } k = \frac{T}{4}$$

Kun $k > \frac{T}{4}$, menee x_2 yli $\frac{T}{2}$:n, joten tilannetta ei käsitellä tässä.

Kuva 4.7. Kun $k < \frac{T}{4}$, oikealla puolella sykliä voi minimierotus olla k yhden kerran — kun $x_3 = k$ — ja vasemmalla puolella kaksi kertaa — kun $x_3 = x_2 + k$ ja $x_3 = x_1 - k$. Yhteensä d_1 :n arvo k esiintyy tilanteessa siis kolme kertaa.

$$\text{fr}_{d_1}(k) = 3 \quad \text{kun } k < \frac{T}{4}$$

$x_2 = 2k + 1, \dots, [\frac{T}{2}]$: ($\neq \frac{T}{2}$) Kuva 4.8. Kun x_2 on yli $2k$ ja alle $\frac{T}{2}$, mahtuu molemmille puolille sykliä 2 kertaa k :n mittainen minimierotus eli yhteensä niiden frekvenssi on neljä. Näissä tilanteissa x_3 :n arvot ovat $x_1 + k$, $x_2 - k$, $x_1 - k$ ja $x_2 + k$. Ehdot $x_2 \leq [\frac{T}{2}]$ ja $x_2 \geq 2k + 1 \Leftrightarrow k \leq \frac{x_2 - 1}{2}$ rajoittavat k :n olemaan korkeintaan $\frac{[\frac{T}{2}] - 1}{2}$.

$$\text{fr}_{d_1}(k) = 4 \quad \text{kun } k \leq \frac{[\frac{T}{2}] - 1}{2}$$

$x_2 = \frac{T}{2}$: (T parillinen) Kuva 4.9. Rajatapaus eli suurin mahdollinen minimierotuksen arvo on $k = \frac{T}{4}$, jonka toteutuminen edellyttää T :n olevan neljällä jaollinen. Se voi esiintyä kahdesti: $x_3 = k = \frac{T}{4}$ ja $x_3 = T - k = T - \frac{T}{4}$. Tätä suurempi ei minimierotus voi tässä tilanteessa olla.

$$\text{fr}_{d_1}(k) = \begin{cases} 2 & \text{kun } k = \frac{T}{4} \\ 0 & \text{kun } k > \frac{T}{4} \end{cases}$$

Jos $k < \frac{T}{4}$, minimierotus k esiintyy neljästi. x_3 :n arvot ovat samat kuin edellisessä kohdassa ($x_2 = 2k + 1, \dots, [\frac{T}{2}]$).

$$\text{fr}_{d_1}(k) = 4 \quad \text{kun } k < \frac{T}{4}$$

4.3.2 d_1 :n frekvensijakauma

Jotta saataisiin selville kokonaisuudessaan $\text{fr}_{d_1}(k)$, on laskettava yhteen ehdolliset frekvenssit $\text{fr}_{d_1}(k \mid x_2)$ yli kaikkien x_2 :ten: $x_2 = 0, \dots, T - 1$. Kun T on pariton, kaikki tapaukset $x_2 = 1, \dots, \lfloor \frac{T}{2} \rfloor$ on laskettava summaan kahdesti, sillä kullakin arvolla on identtinen tapaus, peilikuva, syklin toisella puolella samalla etäisyydellä T :stä eli nolasta. Parilliselle T :lle kohta $x_2 = \frac{T}{2}$ lasketaan summaan vain kerran, koska sillä ei ole peilikuvatapausta.

Ehdolliset d_1 :n frekvenssit x_2 :n sijainnin suhteen riippuvat k :sta. Kaikkiaan eri tilanteita on seuraavasti: (ks. luku 4.3.1)

k	x_2	$\text{fr}_{d_1}(k x_2)$	$\sum_{x_2=0}^{T-1} \text{fr}_{d_1}(k x_2)$	T
$1 \leq k \leq \frac{[\frac{T}{2}]-1}{2}$	$x_2 < k$	0	k	
	$x_2 = k$	$T - 3k + 1$	1	
	$k + 1 \leq x_2 \leq 2k - 1$	2	$k - 1$ *	
	$x_2 = 2k$	3	1	
	$2k + 1 \leq x_2 \leq [\frac{T}{2}]$	4	$[\frac{T}{2}] - 2k$	
	$x_2 = \frac{T}{2}$	4	1	$T \geq 6$
$\frac{[\frac{T}{2}]-1}{2} < k < \frac{T}{4}$ **	$x_2 < k$	0	k	
	$x_2 = k$	$k + 2$	1	
	$k + 1 \leq x_2 \leq 2k - 1$	2	$k - 1$	
	$x_2 = 2k$	3	1	
	$x_2 = \frac{T}{2}$	4	1	$T \geq 5$
	$k = \frac{T}{4}$	$x_2 < k$	0	k
$x_2 = k$		$k + 1$	1	
$k + 1 \leq x_2 \leq 2k - 1$		2	$k - 1$	
$x_2 = 2k = \frac{T}{2}$		2	1	
$\frac{T}{4} < k < \frac{T}{3}$	$x_2 < k$	0	k	
	$x_2 = k$	$T - 3k + 1$	1	
	$k < x_2 < T - 2k$ ***	2	$T - 3k - 1$	
	$x_2 = T - 2k$ ***	1	1	
	$T - 2k + 1 \leq x_2 \leq 2k - 1$	0	- ****	
	$x_2 = \frac{T}{2}$	0	1	$T \geq 7$
$k = \frac{T}{3}$	$x_2 < k$	0	k	
	$x_2 = k$	1	1	
	$k + 1 \leq x_2 \leq 2k - 1$	0	$k - 1$	
	$x_2 = \frac{T}{2}$	0	1	$T \geq 4$

* Jos $k = 1$, tätä riviä ei esiinny.

** Tilanne, missä kokonaislukuarvoinen k osuu välille $[\frac{[\frac{T}{2}]-1}{2}, \frac{T}{4}]$, voi tapahtua vain, kun $T = 4k + 1$. Seuraavissa tarkasteluissa ei siksi käsitellä tilannetta välinä vaan yksittäisenä arvona, $k = \frac{T-1}{4}$.

*** Tämä erikoisjaottelu välillä $k + 1 \leq x_2 \leq 2k - 1$ on tarpeen vain, kun $k \in]\frac{T}{4}, \frac{T}{3}[$.

**** Lukumäärää ei tarvitse saada selville, sillä se kerrotaan joka tapauksessa nolllalla. Koska se lisäksi on vaikea määrittää, sitä ei ole laskettu.

Lasketaan nyt kokonaisfrekvenssit kullekin k :n arvolle yli kaikkien x_2 :n arvojen. T on kiinteä luku. Laskut on tehtävä erikseen parittomalle ja parilliselle T :lle, sillä parillisella T :llä kohta $x_2 = \frac{T}{2}$ tulee mukaan summaan vain kerran ja parittomalla T :llä kaikki kohdat paitsi $x_2 = 0$ lasketaan summaan kahdesti.

$1 \leq k < \frac{T-1}{4}$:

$$\begin{aligned} & \text{fr}_{d_1}(k \mid T \text{ pariton}) \\ &= \text{fr}_{d_1}(k \mid x_2 = 0) + 2(\text{fr}_{d_1}(k \mid x_2 = 1) + \text{fr}_{d_1}(k \mid x_2 = 2) + \cdots + \text{fr}_{d_1}(k \mid x_2 = \lceil \frac{T}{2} \rceil)) \\ &= 0 + 2((T - 3k + 1) + (k - 1)2 + 3 + (\frac{T}{2} - \frac{1}{2} - 2k)4) = 2(3T - 9k) \\ &= 6T - 18k \end{aligned}$$

$$\begin{aligned} & \text{fr}_{d_1}(k \mid T \text{ parillinen}) \\ &= \text{fr}_{d_1}(k \mid x_2 = 0) + 2\left(\text{fr}_{d_1}(k \mid x_2 = 1) + \cdots + \text{fr}_{d_1}(k \mid x_2 = \frac{T}{2} - 1)\right) + \text{fr}_{d_1}(k \mid x_2 = \frac{T}{2}) \\ &= 0 + 2((T - 3k + 1) + (k - 1)2 + 3 + (\frac{T}{2} - (2k + 1))4) + 4 = 2(3T - 9k - 2) + 4 \\ &= 6T - 18k \end{aligned}$$

$k = \frac{T-1}{4}$: Tilannetta, missä T olisi parillinen, ei tarvitse käsitellä, sillä tämä k :n arvo esiintyy vain, kun T on $4k + 1$ eli pariton.

$$\begin{aligned} & \text{fr}_{d_1}(k \mid T \text{ pariton}) \\ &= \text{fr}_{d_1}(k \mid x_2 = 0) + 2\left(\text{fr}_{d_1}(k \mid x_2 = 1) + \cdots + \text{fr}_{d_1}(k \mid x_2 = \lceil \frac{T}{2} \rceil)\right) \\ &= 0 + 2(T - 3k + 1 + (k - 1)2 + 3) + 4 = 2(T - k + 2) = 2T - 2k + 4 \end{aligned}$$

Koska tarkka arvo k :lle on tiedossa, voidaan se vielä sijoittaa lausekkeeseen:

$$2T - 2k + 4 = 2T - 2\frac{T-1}{4} + 4 = \frac{3}{2}T + \frac{9}{2}$$

Tätä k :n arvoa edeltävillä ja seuraavilla k :n arvoilla frekvenssi on $6T - 18k$. Jos tähän lausekkeeseen sijoitetaan $k = \frac{T-1}{4}$, saadaan myös tulos $\text{fr}_{d_1}(k) = \frac{3}{2}T + \frac{9}{2}$. Siispä tätä k :ta ei tarvitse enää huomioida erikoistapauksena.

$k = \frac{T}{4}$: Koska T on jaollinen neljällä, se on aina myös parillinen, joten paritonta tilannetta ei tarvitse käsitellä.

$$\begin{aligned} \text{fr}_{d_1}(k \mid T \text{ parillinen}) &= 0 + 2(T - 3k + 1 + (k - 1)2) + 2 = 2(T - k - 1) + 2 \\ &= 2T - 2k = 2T - 2\left(\frac{T}{4}\right) = \frac{3}{2}T \end{aligned}$$

$\frac{T}{4} < k < \frac{T}{3}$:

$$\begin{aligned} \text{fr}_{d_1}(k \mid T \text{ pariton}) &= 0 + 2(T - 3k + 1 + (T - 3k - 1)2 + 1) = 0 + 6(T - 3k) \\ &= 6T - 18k \\ \text{fr}_{d_1}(k \mid T \text{ parillinen}) &= 0 + 2(T - 3k + 1 + (T - 3k - 1)2 + 1) + 0 \\ &= 6T - 18k \end{aligned}$$

$k = \frac{T}{3}$: $\text{fr}_{d_1}(k \mid x_2 = \frac{T}{2}) = 0$, joten T :n parillisuus ei vaikuta tulokseen.

$$\text{fr}_{d_1}(k) = 0 + 2(1 + 0) = 2$$

Nyt on tiedossa T :n funktiona d_1 :n mahdolliset arvot ja niiden frekvenssit, kun kolme havaintoa saavat kaikki mahdolliset yhdistelmänsä. Kaikissa tilanteissa edellytetään, että $T \geq 4$, ellei muuta mainita. Kolmen havainnon välillä olevan pienimmän erotuksen jakauma on siis

$$\text{fr}_{d_1}(k) = \begin{cases} 3T - 2 & \text{kun } k = 0 \quad * \\ 6T - 18k & \text{kun } 1 \leq k < \frac{T}{4} \\ \frac{3}{2}T & \text{kun } k = \frac{T}{4} \\ 6T - 18k & \text{kun } \frac{T}{4} < k < \frac{T}{3} \\ 2 & \text{kun } k = \frac{T}{3} \text{ ja } T \geq 6. \end{cases} \quad (4.2)$$

* Nollan frekvenssi: ks. kaava 4.1.

4.3.3 Kaikkien havaintokombinaatioiden kokonaismäärä

Jotta saataisiin selville kunkin edellisessä luvussa saadun erotuksen todennäköisyys, on frekvenssit jaettava kaikkien havaintokombinaatioiden kokonaismäärällä. Kolmen havainnon muodostamia kombinaatioita T :n mittaisessa syklissä on T^2 , sillä jokaista x_2 :ta vastaa T kappaletta x_3 :n arvoja ja x_1 voidaan pitää vakiona. Tarkistetaan tämä vielä laskemalla yhteen kaikki k :sta riippuvat frekvenssit kiinteälle T . Lasketaan ensin tilanne, missä T ei ole kolmella eikä neljällä jaollinen, sillä $\frac{T}{3}$:n ja $\frac{T}{4}$:n frekvenssit tuovat

poikkeuksia perusalueen $k \in [1, \frac{T}{3}]$ summalausekkeeseen.

$$\begin{aligned} \sum_{k=0}^{\lfloor \frac{T}{3} \rfloor} \text{fr}_{d_1}(k) &= 3T - 2 + \sum_{k=1}^{\lfloor \frac{T}{3} \rfloor} (6T - 18k), \text{ kun } T \text{ ei jaoll. 3:lla eikä 4:llä} \\ &= 3T - 2 + \lfloor \frac{T}{3} \rfloor 6T - 9 \lfloor \frac{T}{3} \rfloor (\lfloor \frac{T}{3} \rfloor + 1) = 3T - 2 + \lfloor \frac{T}{3} \rfloor (6T - 9(\lfloor \frac{T}{3} \rfloor + 1)) \\ &= \begin{cases} 3T - 2 + \frac{T-1}{3} (6T - 9(\frac{T-1}{3} + 1)) = T^2 & \text{kun } T = 3a + 1, a \in \mathbb{N} \\ 3T - 2 + \frac{T-2}{3} (6T - 9(\frac{T-2}{3} + 1)) = T^2 & \text{kun } T = 3a + 2 \end{cases} \end{aligned}$$

Laskun loppuosa on jaettu kahteen tilanteeseen: toisessa T on muotoa $3a + 1$, kun a on jokin luonnollinen luku, ja toisessa T on muotoa $3a + 2$. Koska molemmista vaihtoehdoista päädyttiin samaan tulokseen, on tulos yksikäsitteinen kolmella ja neljällä jaottomalle T :lle. Lasketaan seuraavaksi kolmella jaollisen T :n tilanne olettaen, että T ei ole neljällä jaollinen:

$$\begin{aligned} \sum_{k=0}^{\frac{T}{3}-1} \text{fr}_{d_1}(k) &= \text{fr}_{d_1}(0) + \sum_{k=1}^{\frac{T}{3}-1} \text{fr}_{d_1}(k) + \text{fr}_{d_1}(\frac{T}{3}), \text{ kun } T \text{ jaoll. 3:lla mutta ei 4:llä} \\ &= 3T - 2 + \sum_{k=1}^{\frac{T}{3}-1} (6T - 18k) + 2 \\ &= 3T - 2 + (\frac{T}{3} - 1)6T - 9\frac{T}{3}(\frac{T}{3} - 1) + 2 = T^2. \end{aligned}$$

Tulos oli myös tässä tapauksessa odotettu T^2 . Tutkitaan vielä tilanne, jossa T on jaollinen neljällä mutta ei kolmella. Vähennetään alkuperäisen laskun summalausekkeesta termi, jossa $k = \frac{T}{4}$ ja korvataan se lisäämällä summaan oikea lauseke $\frac{3}{2}T$:

$$\begin{aligned} \sum_{k=0}^{\lfloor \frac{T}{3} \rfloor} \text{fr}_{d_1}(k) &= 3T - 2 + \sum_{k=1}^{\lfloor \frac{T}{3} \rfloor} (6T - 18k) - (6T - 18\frac{T}{4}) + \frac{3}{2}T \\ &= 3T - 2 + \sum_{k=1}^{\lfloor \frac{T}{3} \rfloor} (6T - 18k) - \overbrace{\frac{3}{2}T + \frac{3}{2}T}^{=0}. \end{aligned}$$

Neljällä jaollisen T :n aiheuttamat muutokset laskussa ovat siis yhteensä nolla, joten se ei muuta edellä saatuja tuloksia, vaikka T olisikin jaollinen myös kolmella.

Nyt siis tiedetään, kuinka monta eri tilannetta syklin kehällä yhteensä oli, kun laskettiin kunkin eri erotuksen d_1 frekvenssit. Täten kunkin erotuksen k todennäköisyys kaikkien mahdollisten d_1 :n arvojen joukossa saadaan selville jakamalla sen frekvenssi kokonaisfrekvenssillä T^2 .

4.4 Toiseksi pienimmän erotuksen d_2 jakauma

Nimetään kolmen havainnon välillä olevista kolmesta erotuksesta toiseksi pienin d_2 :ksi ja lasketaan myös sille todennäköisyysjakauma riippumattomassa tilanteessa. Menetelmä on selitetty luvussa 4.2.

4.4.1 Nollan ja ykkösen mittaisten erotusten frekvenssit

Lasketaan ensin esimerkin vuoksi nollan ja ykkösen mittaisten erotusten esiintymisfrekvenssit. Toiseksi suurin erotus voi olla nolla vain, kun kaikki kolme havaintoa ovat samassa pisteessä. Tämä tapahtuu yhden kerran kaikkien mahdollisten havaintokombinaatioiden joukossa, kun x_1 on kiinnitetty. Tällöin kaikki kolme havaintoa ovat samassa pisteessä.

Ykkösen pituisen erotuksen frekvenssiä laskettaessa on tutkittava erikseen monta tilannetta x_2 :n sijainnin suhteen:

$x_2 = 0$: Kuva 4.10. Erotus $d_2 = 1$ esiintyy kahdesti: kun $x_3 = 1$ ja $x_3 = T - 1$ (yksi askel syklin nollakohdasta vastapäivään).

$x_2 = 1$: Erotus $d_2 = 1$ esiintyy neljästi: x_3 :n arvot ovat tällöin peräkkäin välillä $[T - 1, 2]$.

$x_2 = 2$: Erotus $d_2 = 1$ esiintyy kerran: kun x_3 on x_1 :n ja x_2 :n puolivälissä eli kohdassa yksi. Muulloin toiseksi suurin erotus on vähintään kaksi.

$x_2 \geq 3$: Maksimierotus d_2 on väistämättä vähintään kaksi, $\text{fr}_{d_2}(1) = 0$.

Kun lasketaan yhteen nämä kaikki ykkösen frekvenssin eri tapaukset, saadaan yhteensä 12, kun $T \geq 5$: $\text{fr}_{d_2}(1) = 2 + 2(4 + 1 + 0) = 12$. Kun $T = 4$, $\text{fr}_{d_2}(1) = 2 + 2 \cdot 4 + 2 = 12$ sillä tapausta $x_2 = 2$ ei lasketa kokonaissummaan kuin kerran. Tämä on kuitenkin sama tulos kuin T :n ollessa vähintään viisi, joten tulos pätee, kun $T \geq 4$. Nollan ja ykkösen frekvenssit ovat siis

$$\text{fr}_{d_2}(0) = 1$$

$$\text{fr}_{d_2}(1) = 12.$$

4.4.2 d_2 :n arvojen frekvenssien laskeminen ja todennäköisyysjakauma

Määritetään sitten yleisessä tilanteessa minkä tahansa k :n frekvenssi siten, että annetaan x_2 :n saada vuorollaan kaikki arvot välillä $[0, \frac{T}{2}]$ ja lasketaan valitun erotuksen esiintymiset kaikissa näissä tilanteissa. Tarkkoja tilannekuvauksia ei enää esitetä vaan tulokset ilmoitetaan seuraavassa taulukossa.

Sarakkeessa $\text{fr}_{d_2}(k \mid x_2)$ on ilmoitettu kyseisen erotuksen k esiintymisfrekvenssi yhdellä kierroksella, jonka aikana x_2 on vakio ja x_3 saa peräkkäin kaikki arvot syklin kehällä. Viite-sarakkeessa on mahdollinen viite tilannetta havainnollistavaan kuvaan tai suluissa viite taulukon jälkeen olevaan selitykseen. Syklin pituus T pidetään kiinteänä ja k ilmoitetaan T :n funktiona.

k	x_2	$\text{fr}_{d_2}(k \mid x_2)$	Viite	$\sum_{x_2=0}^{T-1} \text{fr}_{d_2}(k \mid x_2)$	T
0	0	1			
	$1, \dots, \lfloor \frac{T}{2} \rfloor$	0		1	$T \geq 4$
1	0	2			
	1	4			
	2	1			
	$3, \dots, \lfloor \frac{T}{2} \rfloor$	0		12	$T \geq 4$
$[2, \frac{T}{4}[$	$0, \dots, k-1$	2	4.11		
	k	$2k+2$	4.12		
	$k+1, \dots, 2k-1$	2	4.13		
	$2k$	1			
	$2k+1, \dots, \lfloor \frac{T}{2} \rfloor$	0	(1)	$12k$	$T \geq 9$
$\frac{T}{4}$	$0, \dots, k-1$	2			
	k	$2k+2$			
	$k+1, \dots, \frac{T}{2}-1$	2	4.14		
	$\frac{T}{2}$	2	4.15	$12k$	$T \geq 8$
$]\frac{T}{4}, \frac{T}{3}[$	$0, \dots, k-1$	2			
	k	$2k+2$			
	$k+1, \dots, T-2k-1$	2	(2)		
	$T-2k$	3	4.16		
	$T-2k+1, \dots, \lfloor \frac{T}{2} \rfloor$	4	(3)	$12k$	$T \geq 10$
$\frac{T}{3}$	$0, \dots, k-1$	2			
	k	$2k+1$	4.17		
	$k+1, \dots, \lfloor \frac{T}{2} \rfloor$	4		$12k-4$	$T \geq 9$
$]\frac{T}{3}, \frac{T}{2}[$	$0, \dots, T-2k-1$	2			
	$T-2k$	1			
	$T-2k+1, \dots, k-1$	0	(4)		
	k	$2(T-2k+1)$	(5)		
	$k+1, \dots, \lfloor \frac{T}{2} \rfloor$	4		$12T-24k$	$T \geq 9$
$\frac{T}{2}$	0	1			
	$1, \dots, \frac{T}{2}-1$	0			
	$\frac{T}{2}$	2		3	$T \geq 4$

- (1) Jos $k = \frac{T}{3}$ ja $T = 3k + 1$, tätä tilannetta ei esiinny, sillä $k + 1$ eli alueen alaraja on suurempi kuin yläraja $T - 2k - 1$.
- (2) Jos $k = [\frac{T}{3}]$ ja $T = 3k + 1$, tätä tilannetta ei esiinny, sillä $k + 1 = T - 2k$ eli alueen alaraja on suurempi kuin yläraja.
- (3) Jos $k = \frac{T}{4} + \frac{1}{4}$, tätä tilannetta ei esiinny, sillä alueen alaraja on suurempi kuin yläraja. Jos taas $k < \frac{T+6}{4}$, rivi esiintyy kerran, vaikka alaraja on suurempi kuin yläraja.
- (4) Jos $k < \frac{T}{3} + \frac{2}{3}$, tätä riviä ei esiinny.
- (5) $\text{fr}_{d_2}(k) = (T - k + 1) - (2k - (T - k + 1)) = 2(T - 2k + 1)$

Lasketaan sitten frekvenssien summa yli kaikkien x_2 :n vaihtoehtojen kullekin k :lle. Nämä tulokset ovat näkyvillä myös taulukon sarakkeessa $\sum_{x_2=0}^{T-1} \text{fr}_{d_2}(k | x_2)$.

$2 \leq k < \frac{T}{4}$: $\text{fr}_{d_2}(k) = 2 + 2((k-1)2 + 2k + 2 + (2k-1-k)2 + 1 + 0) = 2 + 2(4k + 2k - 1) = 12k$
 T :n parillisuutta ei otettu huomioon, sillä x_2 :n kohdassa $\frac{T}{2}$, missä sillä ainoastaan olisi vaikutusta, k :n frekvenssi on nolla.

$$k = \frac{T}{4} : \text{fr}_{d_2}(\frac{T}{4}) = 2 + 2(2(k-1) + 2k + 2 + 2(\frac{T}{2} - 1 - k)) + 2 = 2 + 2(T + 2k - 2) + 2 = 2T + 4k = 12k, \text{ kun sijoitetaan } k = \frac{T}{4}.$$

$$\begin{aligned} \frac{T}{4} < k < \frac{T}{3} : \text{fr}_{d_2}(k | T \text{ pariton}) \\ &= 2 + 2((k-1)2 + 2k + 2 + (T-2k-1-k)2 + 3 + ([\frac{T}{2}] - (T-2k))4) \\ &= 2 + 2(4k + (T-3k-1)2 + 3 + 8k - 4T + 4\frac{T-1}{2}) = 2 + 2(6k - 1) = 12k \end{aligned}$$

$$\begin{aligned} \text{fr}_{d_2}(k | T \text{ parillinen}) \\ &= 2 + 2((k-1)2 + 2k + 2 + (T-2k-1-k)2 + 3 + (\frac{T}{2} - 1 - (T-2k))4) + 4 \\ &= 12k \end{aligned}$$

$$k = \frac{T}{3} : \text{fr}_{d_2}(\frac{T}{3}) = 2 + 2((k-1)2 + 2k + 1 + ([\frac{T}{2}] - k)4) = 12k - 4 \text{ (sij. } k = \frac{T}{3}) = 12\frac{T}{3} - 4 = 4T - 4 \text{ kun } T \text{ on pariton. Lasku on vastaava myös parilliselle } T\text{:lle.}$$

$$\begin{aligned} \frac{T}{3} < k < \frac{T}{2} : \text{Kun } T \text{ on pariton, suurin } x_2\text{:n saama arvo on } \frac{T-1}{2}, \text{ jota on käytetty laskussa.} \\ \text{fr}_{d_2}(k | T \text{ pariton}) &= 2 + 2((T-2k-1)2 + 1 + 2(T-2k+1) + (\frac{T-1}{2} - k)4) \\ &= 2 + 2(4T - 8k + 1 + 2T - 4k - 2) = 12T - 24k \end{aligned}$$

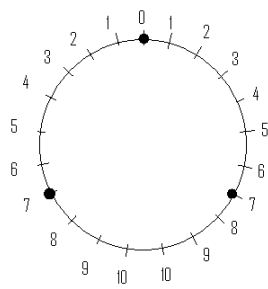
$$\begin{aligned} \text{Kun } T \text{ on parillinen, } x_2\text{:n suurin arvo on } k = \frac{T}{2} \\ \text{fr}_{d_2}(k | T \text{ parillinen}) &= 2 + 2((T-2k-1)2 + 1 + 2(T-2k+1) + (\frac{T}{2} - 1 - k)4) + 4 \\ &= 2 + 2(4T - 8k + 1 + 2T - 4k - 4) + 4 = 12T - 24k \end{aligned}$$

$$k = \frac{T}{2} : \text{fr}_{d_2}\left(\frac{T}{2}\right) = 1 + 2 = 3$$

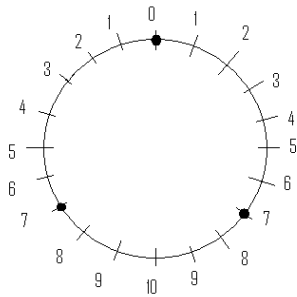
Nyt voidaan siis koota yhteen d_2 :n eri arvojen frekvenssit, kun kolme havaintoa saavat kaikki kombinaatiot syklin kehällä. Kaavojen oikeellisuus edellyttää tarpeeksi suurta T :tä (ks. edellinen taulukko). Toiseksi pienimmän erotuksen frekvenssijakauma on siis

$$\text{fr}_{d_2}(k) = \begin{cases} 1 & \text{kun } k = 0 \\ 12k & \text{kun } 1 \leq k < \frac{T}{3} \\ 4T - 4 & \text{kun } k = \frac{T}{3} \text{ ja } k \geq 2 \\ 12T - 24k & \text{kun } \left[\frac{T}{3}\right] + 1 \leq k < \frac{T}{2} \\ 2 & \text{kun } k = \frac{T}{2}. \end{cases} \quad (4.3)$$

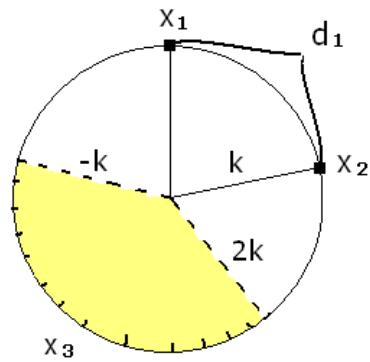
d_2 :n pistetodennäköisyysjakauma saadaan, kun kukin frekvenssi jaetaan kombinaatioiden kokonaismäärällä T^2 (ks. luku 4.3.3).



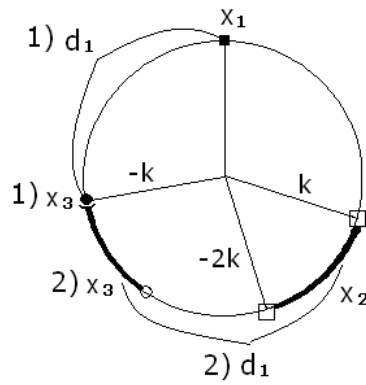
Kuva 4.1: Epäsynchronian maksimi parittomassa syklissä.



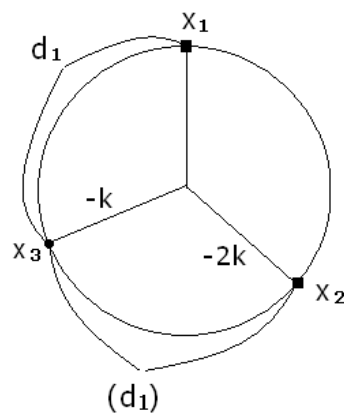
Kuva 4.2: Epäsynchronian maksimi parillisessa syklissä.



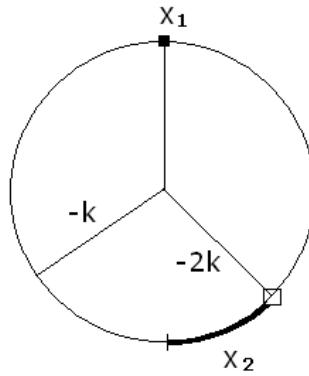
Kuva 4.3: $k < \frac{T}{3}$ ja $x_2 = k$



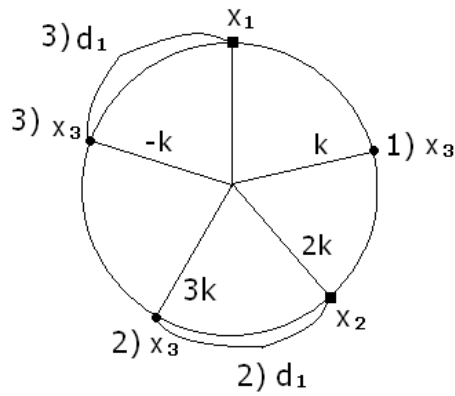
Kuva 4.4: $\text{fr}_{d_1} = 2$ kun $k < \frac{T}{3}$ ja $x_2 = k + 1, \dots, T - 2k - 1$.



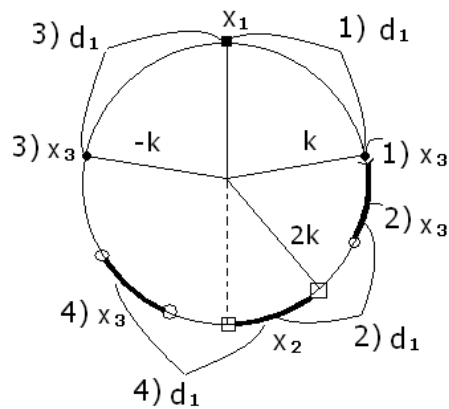
Kuva 4.5: $\text{fr}_{d_1} = 1$ kun $k < \frac{T}{3}$ ja $x_2 = T - 2k$



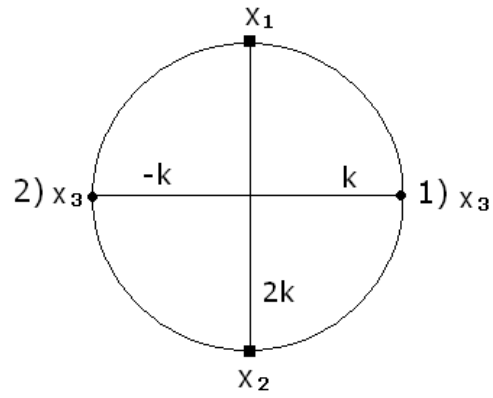
Kuva 4.6: $\text{fr}_{d_1} = 0$ kun $k < \frac{T}{3}$ ja $x_2 = T - 2k + 1, \dots, 2k - 1$



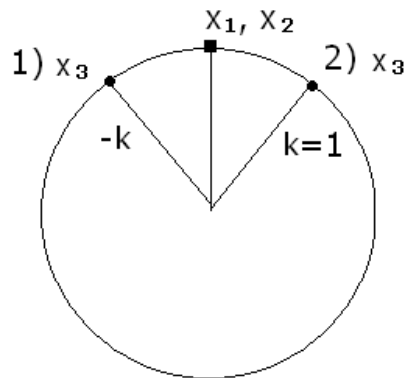
Kuva 4.7: $\text{fr}_{d_1} = 3$ kun $k < \frac{T}{4}$ ja $x_2 = 2k$



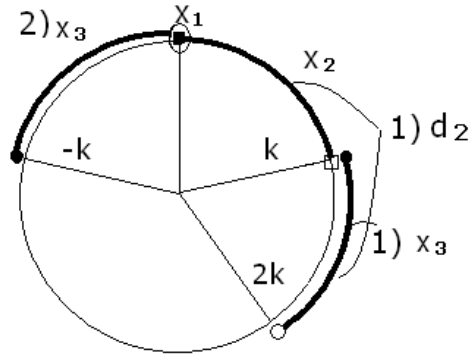
Kuva 4.8: $\text{fr}_{d_1} = 4$ kun $k \leq \frac{\lfloor T/2 \rfloor - 1}{2}$ ja $x_2 = 2k + 1, \dots, \lfloor \frac{T}{2} \rfloor$. Näissä tilanteissa x_3 :n arvot ovat (1) k (2) $x_2 - k$ (3) $x_1 - k$ ja (4) $x_2 + k$.



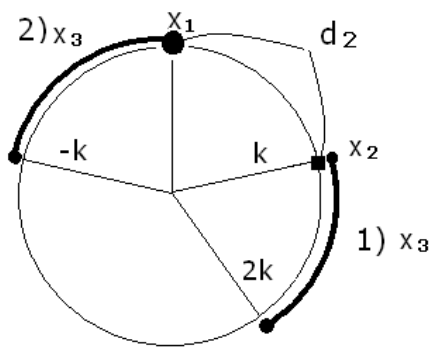
Kuva 4.9: $\text{fr}_{d_1} = 2$ kun $k = \frac{T}{4}$ ja $x_2 = \frac{T}{2}$



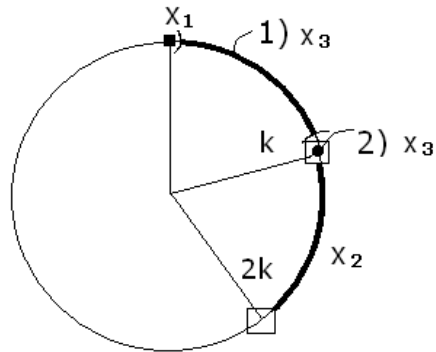
Kuva 4.10: $\text{fr}_{d_2}(1) = 2$ kun $x_2 = x_1$



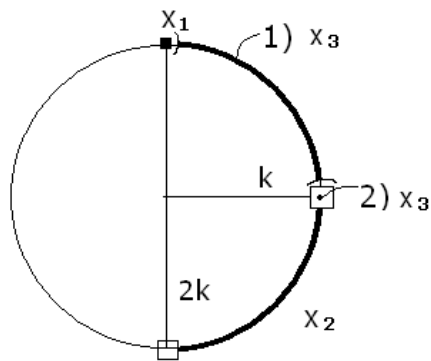
Kuva 4.11: $\text{fr}_{d_2}(k) = 2$ kun $k \in [2, \frac{T}{4}[$ ja $x_2 = 0, \dots, k - 1$. Tällöin $x_3 = x_2 + k$ tai $x_3 = x_1 - k$.



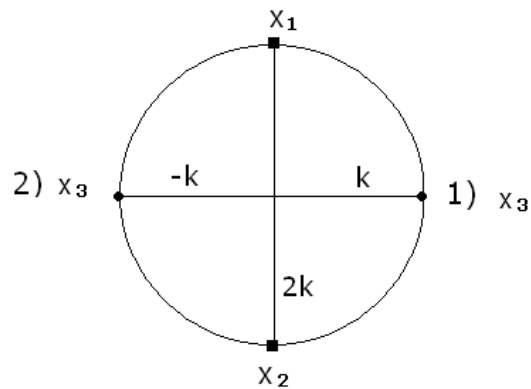
Kuva 4.12: $\text{fr}_{d_2}(k) = 2k + 2$ kun $k \in [2, \frac{T}{4}[$ ja $x_2 = k$, sillä $d_2 = k$ kun $x_3 \in [x_2, x_2 + k]$ tai $x_3 \in [x_1 - k, x_1]$, ja näiden välien yhteenlaskettu pituus on $(k + 1) + (k + 1) = 2k + 2$.



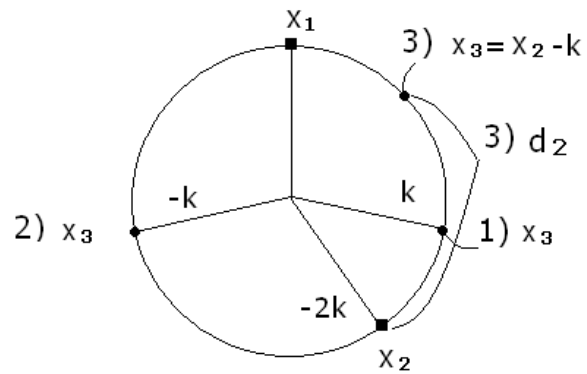
Kuva 4.13: $\text{fr}_{d_2}(k) = 2$ kun $k \in [2, \frac{T}{4}[$ ja $x_2 = k + 1, \dots, 2k - 1$. x_3 on tällöin $x_2 - k$ tai $x_1 + k = k$.



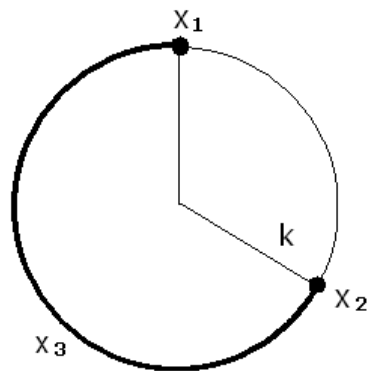
Kuva 4.14: $\text{fr}_{d_2}(\frac{T}{4}) = 2$ kun $x_2 = k + 1, \dots, \frac{T}{2} - 1$. x_3 on tällöin $x_2 - k$ tai $x_1 + k = k$.



Kuva 4.15: $\text{fr}_{d_2}(\frac{T}{4}) = 2$ kun $x_2 = \frac{T}{2}$



Kuva 4.16: $\text{fr}_{d_2}(k) = 3$ kun $k \in [\frac{T}{4}, \frac{T}{3}[$ ja $x_2 = T - 2k$



Kuva 4.17: $\text{fr}_{d_2}(\frac{T}{3}) = 2k + 1$ kun $x_2 = k$. Tällöin $x_3 \in [x_2, x_1]$ ja tämän välin pituus on $T - (k - 1) = 3k - k + 1 = 2k + 1$.

5 Simuloinnin tuottama erotusten jakauma kolmen havainnon otoksille

Tutkitaan kolmen ja neljän havainnon otoksissa kahden ja kolmen pienimmän erotuksen jakaumaa simuloitujen aineistojen avulla. Kolmen havainnon otoksien suhteen voidaan vertailla saatuja erotusjakaumia teoreettisiin jakaumiin. Toteutetaan simulointi R-ohjelmalle [6] tehdyllä funktiolla, koodi on liitteessä 2. Menetelmä pohjautuu tasajakautuneiden kokonaislukujen generointiin välille $[0, T - 1]$, satunnainen valinta toteutetaan funktiolla `sample(0:(T-1), 1, replace=TRUE)`. Kolmesta tai neljästä tällaisesta havainnosta kootaan otos ja n otoksesta havaintoaineisto. Aineistoja tuotetaan N kpl. Otetaan kunkin simulointikierron tuloksena talteen syklistä ajankäsitystä käyttäen kaikkien n otoksen sisältämät erotukset d_1 , d_2 ja mahdollisesti d_3 .

5.1 Simulointi kolmen havainnon otoksille ja tulosten vertailu teoriaan

Simuloidaan $n=10$ otoksen kokoisia aineistoja, sillä oikea havaintoaineisto sisältää kymmenen kolmesta havainnosta koostuvaa otosta, ja toistetaan aineiston simulointi $N=50\,000$ kertaa. Valitaan syklin maksimiksi $T=21$, kuten oletetaan havaintoaineistossa olevan. Lasketaan d_1 :lle ja d_2 :lle kunkin mahdollisen erotuksen arvon k osuus joka aineistossa ($k_1 = 0, \dots, 7$ ja $k_2 = 0, \dots, 10$) ja keskiarvo kyseisistä osuuksista yli simulointikierrosten. Näin saadaan empiiriset erotusten d_1 ja d_2 pistetodennäköisyysjakaumat, joita voidaan verrata vastaaviin teoreettisiin jakaumiin, jotka on esitetty kaavoissa 4.2 ja 4.3. Tapauksessa $T = 21$ pienimmän erotuksen frekvenssijakauma on

$$\begin{aligned} \text{fr}_{d_1}(k) &= \begin{cases} 3T - 2 & \text{kun } k = 0 \\ 6T - 18k & \text{kun } 1 \leq k < \frac{T}{4} \\ \frac{3}{2}T & \text{kun } k = \frac{T}{4} \\ 6T - 18k & \text{kun } \frac{T}{4} < k < \frac{T}{3} \\ 2 & \text{kun } k = \frac{T}{3} \text{ ja } T \geq 6 \end{cases} \\ &= \begin{cases} 61 & \text{kun } k = 0 \\ 126 - 18k & \text{kun } 1 \leq k \leq 6 \\ 2 & \text{kun } k = 7 \end{cases} \end{aligned}$$

ja toiseksi pienimmän erotuksen frekvenssijakauma on

$$\begin{aligned} \text{fr}_{d_2}(k) &= \begin{cases} 1 & \text{kun } k = 0 \\ 12k & \text{kun } 1 \leq k < \frac{T}{3} \\ 4T - 4 & \text{kun } k = \frac{T}{3} \text{ ja } k \geq 2 \\ 12T - 24k & \text{kun } \lceil \frac{T}{3} \rceil + 1 \leq k < \frac{T}{2} \\ 2 & \text{kun } k = \frac{T}{2} \end{cases} \\ &= \begin{cases} 1 & \text{kun } k = 0 \\ 12k & \text{kun } 1 \leq k \leq 6 \\ 80 & \text{kun } k = 7 \\ 252 - 24k & \text{kun } 8 \leq k \leq 10. \end{cases} \end{aligned}$$

Kun odotetut frekvenssit jaetaan $T^2 = 441$:llä, saadaan todennäköisyydet kullekin arvolle. Esitetään niistä likiarvot taulukossa 5.1.

Taulukko 5.1: Kolmen havainnon jakauman pistetodennäköisyydet kullekin osuudelle $k = 0, \dots, 10$, kun $T = 21$.

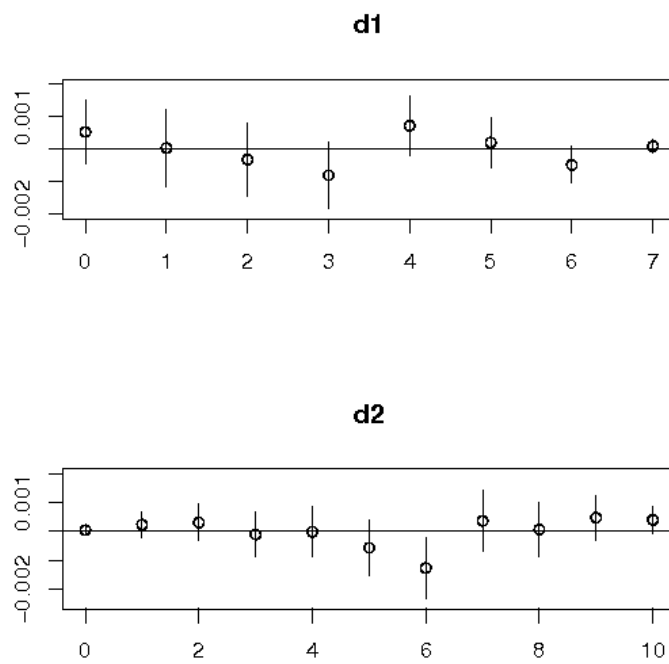
	0	1	2	3	4	5	6	7	8	9	10
d_1	0.138	0.245	0.204	0.163	0.122	0.082	0.041	0.005			
d_2	0.002	0.027	0.054	0.082	0.109	0.136	0.163	0.181	0.136	0.082	0.027

Kuvassa 5.1 on esitetty erotusten d_1 ja d_2 kunkin mahdollisen arvon k osuuden erotus teoreettisen ja simuloitun tuloksen välillä. Lisäksi on laskettu tämän eron 95 %:n luottamusväli simuloinnin pohjalta ja katsottu, sisältyykö nolla tälle välille, piirtämällä nollan korkeudelle viiva. Luottamusväli on laskettu käyttäen normaalijakauman kertymäfunktioista saatua arvoa $\Phi^{-1}(0.95)$ välin pituuden määrittämiseen. Tällöin oletetaan simulointituloksesta saatavan erotuksen arvon k osuuden noudattavan normaalijakaumaa. Oletus on perusteltu, koska simulointituloksesta on N tuloksen keskiarvo ja tässä $N = 50000$, siis suuri. Kun tietyn erotuksen osuus otoksissa on keskimäärin $\overline{P(k)_{sim}}$ ja erotus teoreettisen ja simuloitun tuloksen välillä on $P(k)_{teor} - \overline{P(k)_{sim}} = \delta$, on luottamusväli

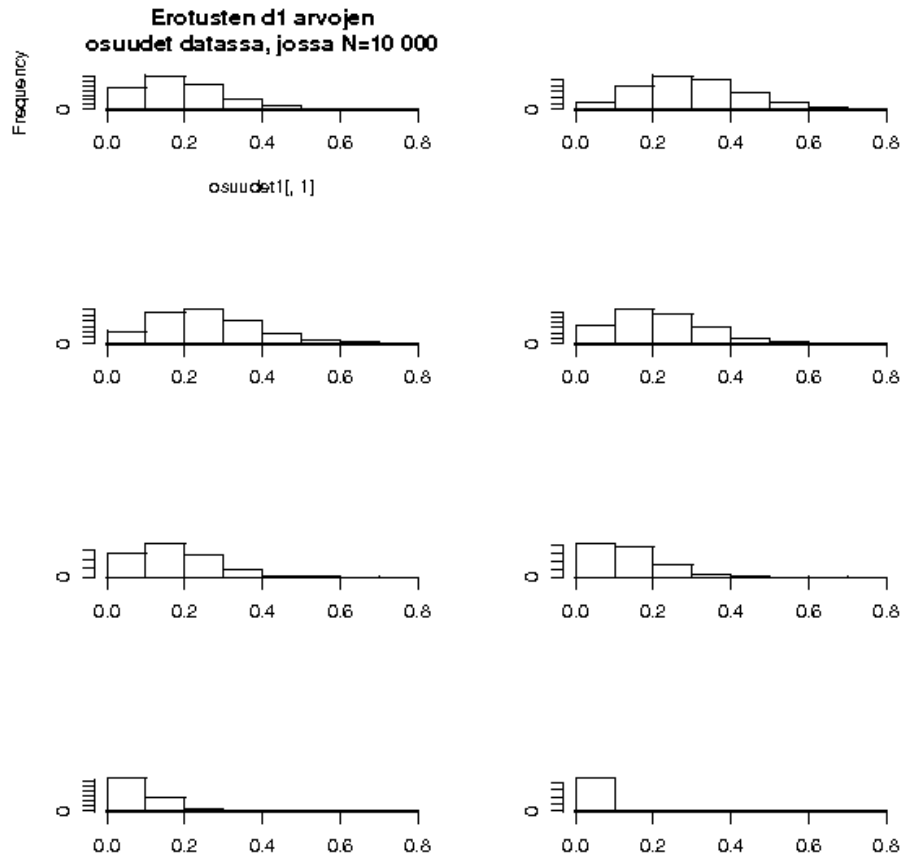
$$\delta \pm 1.959964 \cdot \sqrt{\frac{\text{Var}(\overline{P(k)_{sim}})}{N}}.$$

5.2 Simulointimallin luotettavuus

Simulointituloksien 95 %:n luottamusvälit sisältävät teoreettiset tulokset muissa 18 tilanteessa paitsi kohdassa $d_2 = 6$. Teorian mukaan 95 %:n luottamustasolla nollassa nollahypoteesin mukainen aineisto ei sisälly luottamusvälille 5 %:ssa testatuista tapauksista. Nyt



Kuva 5.1: d_1 :n ja d_2 :n teoreettisen ja simuloidun pistetodennäköisyysjakaumien erotuksen keskiarvot ja 95 %:n luottamusvälit



Kuva 5.2: Vasemmalta oikealle d_1 :n osuudet 0:sta 7:ään, jakaumat simuloitun aineiston perusteella. $T = 21$

luottamusvälejä on 19, ja $\frac{1}{19}$ vastaa 5 %:a. Koska kyseisessä tilanteessa 99 %:n luottamusväli (ei esitetty kuvassa) kuitenkin sisältää nollan, tilanne ei ole hälyttävä. Lisäksi kuvissa ei ole havaittavissa systemaattisia ilmiöitä ja keskiarvot (pallot viivan päällä) näyttävät noudattavan normaalijakaumaa niin kuin kuuluukin. Voidaan siis sanoa, että simulointimallin tulokset vastaavat teoriaa.

Kuvasta 5.2 (simulointikoe on eri kuin muissa tuloksissa) näkee, että joidenkin osuuksien k jakaumat ovat oikealle vinoja, koska osuuden odotusarvo on lähellä nollaa. Osuus ei tällöin voi olla juurikaan odotettua pienempi mutta paljon suurempi kylläkin. Vinoissa jakaumissa keskiarvo ei ole paras tunnusluku kuvaamaan jakauman huippua, mutta koska vinous esiintyy samanlaisena sekä simuloituissa aineistossa että oletetussa reaali maailmassa, voidaan keskiarvoa myöhemmin käyttää simuloitun ja todellisen datan vertailussa. Simulointimallia voidaan siis pitää edelleen hyvänä. Vinous lienee myös syynä tilanteeseen $fr_{d_2}(6)$, jossa simuloitun datan keskiarvo osuudelle oli suurem-

pi kuin teorian mukainen eikä mahtunut 95 %:n luottamusvälille. Suuret poikkeavat arvot saivat keskiarvon nousemaan liian korkeaksi suhteessa odotusarvoon.

6 Neljän havainnon välisten erotusten jakauma

Kun neljän havainnon välisiä erotuksia halutaan tutkia, on otettava kuudesta pareittaisesta erotuksesta tutkittavaksi vain jotkin kolme, jotka eivät ole toistensa summia. Tämä voidaan sopia tehtäväksi eri tavoin, kuten on kerrottu luvussa 4.

Tehdään nyt vastaavasti kuten teoriassa aiemmin kolmen havainnon otosten suhteen eli valitaan tutkittavaksi pienimmät mahdolliset erotukset. Neljän havainnon välillä voi olla kolme erotusta, joista mikään ei ole toistensa summa. Tilanne on työläs laskea teoreettisesti, joten tuotetaan erotusten jakauma pelkästään simuloinnin avulla.

6.1 Erotusten d_1 , d_2 ja d_3 jakauma simuloimalla

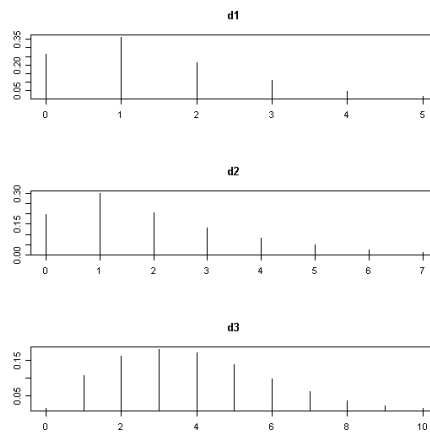
Käytetään neljän havainnon otoksille vastaavaa simulointimallia kuin kolmen havainnon otoksille. Käytetään edelleen syklin maksimina $T=21$:tä mutta valitaan otosten määräksi aineistossa $n=4$ ja simulointikierrosten määräksi $N=100\,000$.

Taulukossa 6.1 on kuvattu erotusten d_1 , d_2 ja d_3 simulointituloksista laskettuja pistetodennäköisyysjakaumia. Jotkin jakaumat ovat vinoja, jolloin keskiarvo ei välttämättä ole paras estimaattori kuvaamaan jakauman huippua mutta käytetään sitä kuitenkin sen tavallisuuden vuoksi.

Kuvassa 6.1 on kaikkien erotusten d_1 , d_2 ja d_3 pistetodennäköisyysjakaumat. Kaikkien osuuksien luottamusvälit ovat niin pienet, että ne eivät erottuisi kuvassa, joten niitä ei ole piirretty.

k	$P(d_1 = k)$	$P(d_2 = k)$	$P(d_3 = k)$
0	0.2633	0.01556	0.008875
1	0.3754	0.13422	0.05113
2	0.2196	0.2168	0.09777
3	0.10526	0.2442	0.13944
4	0.03389	0.2158	0.1758
5	0.002485	0.13431	0.2069
6		0.03785	0.1929
7		0.001333	0.07962
8			0.03332
9			0.01225
10			0.002015

Taulukko 6.1: Simuloiden saadut pistetodennäköisyydet neljän havainnon välisille kolmelle pienimmälle erotukselle



Kuva 6.1: Neljän havainnon välisten kolmen pienimmän erotuksen pistetodennäköisyysjakaumat

7 Empiirinen tutkimus

7.1 Aineisto

Tutkittava aineisto koostuu luonnossa lähekkäin elävien metsämyyränaaraiden synnytyspäivistä. Aineisto on kerätty Konneveden kunnan alueella Keski-Suomessa ($60^{\circ}37'N$, $26^{\circ}20'E$) vuosien 1995–2005 aikana. Myyriä pyydettiin kolme kertaa lisääntymiskauden aikana, toukokuun lopussa, heinäkuun alussa ja elokuun lopussa. Pyynti suoritettiin 20 erillisessä paikassa. Kullakin paikalla oli 4 elävänä pyytävää Ugglan-loukkua. Loukut oli asetettu neliöön 10–15 metrin etäisyydellä toisistaan. Kaikki samalta paikalta saadut naaraat siis elivät lähekkäisillä reviireillä. Loukut viritettiin illalla ja ne tarkastettiin kahtena seuraavana aamuna. Kaikki kiinni saadut naaraat vietiin laboratorioon synnytystä odottamaan. Naarailta otettiin muistiin synnytyspäivä siitä päivästä lukien, jolloin loukut koettiin ensimmäisen kerran. Näin voidaan vertailla lähekkäin eläneiden naaraiden synnytyspäivien välistä aikaerotusta.

Aineistossa (liite 1) on $44 = n_2$ kahden, $10 = n_3$ kolmen ja $6 = n_4$ neljän myyrän otosta. Koska tutkittavana ovat erotukset, pidetään tutkimusten lähtökohtana synnytyspäivien välisiä erotuksia, tarkkoja synnytyspäiviä ei tarkastella. Metsämyyrien tiineysaika on keskimäärin 20 päivää, joten erotus voi olla korkeintaan tämän verran. Lisäksi oletetaan myyrien tulevan aina tiineeksi synnytyksen jälkeisenä päivänä, joten käytetään syklistä ajankäsitystä T :n arvolla 21. Näin asetetaan erotukset 21 ja 0 samankarvoisiksi. Vastaavasti on oletettu useissa muissa vastaavissa tutkimuksissa ([3], [4]). Synnytysten synkroniaa tutkittaessa tarkalla syklin pituudella ei tässä tutkimuksessa ole kuitenkaan merkitystä, sillä lähellä 21:tä olevia erotuksia ei aineistossa ole.

7.2 Synkronian määrittely

Määritellään synnytysten synkronia tarkoittamaan sitä, että synnytykset tapahtuvat samana päivänä. Myös muutaman päivän erotuksia voitaisiin jossain tilanteissa pitää synkronian merkinä, mutta tässä tutkimuksessa keskitytään vain nollaerotuksiin, jotta tulokset olisivat paremmin verrattavissa kahden käytetyn synkroniatyyppin välillä. Toinen tutkittava ilmiö on kiimasykliin synkronia. Metsämyyränaarilla on kiima keskimäärin neljän päivän välein ja muulloin ne eivät parittele. Jos siis halutaan tutkia, ovatko myyrien kiimasyklit synkronoituneita, voidaan mikä tahansa neljällä jaollinen

erotus synnytyspäivien välillä tulkita kiimasykli synkroniaksi. Tämä perustuu siihen, että jos oletetaan naaraiden olleen tiineenä yhtä monta päivää, ne ovat olleet kiimassa samana päivänä, jolloin myös kiimasyklit ovat rinnakkain kulkevia.

Kiimasykli synkroniaa tutkittaessa on kuitenkin luotettavampaa olla käyttämättä syklisen ajan menetelmiä, sillä oletettu tiineysaika, 20 päivää, on vain keskiarvo eivätkä myyrät välttämättä myöskään parittele heti synnyttyään. Niinpä esim. 21 päivän välein synnyttäneistä myyristä ensin synnyttänyt ei välttämättä ole tullut uudelleen tiineeksi samana päivänä kuin toiseksi synnyttänyt on tullut. Tämän vuoksi ei kannata vertailla erotuksia, jotka todellisuudessa ovat kovin pitkiä, vaikka niillä olisikin lyhyt väli oletettuun toisen myyrän synnytyspäivään. Synnytysten synkronian suhteen syklinen aika on järkevä, erityisesti jos myös yhden päivän mittaiset erotukset laskettaisiin synkronian merkiksi. (Kuten todettua, tässä aineistossa ei ole kuitenkaan lainkaan lähellä 20:tä olevia erotuksia.) Neljän päivän pituisen syklin perustaminen synnytysten jatkuvuuteen on huomattavasti riskialttiimpaa, koska kaikkien yli 10 päivän erotusten tulkitseminen neljällä jaolliseksi muuttuu, kun havainnot sijoitetaan sykliin. 21 päivän pituiseen sykliin sijoitettuna neljällä jaolliseksi tulevat nimittäin alunperin 13 ja 17 päivän pituiset erotukset. Näin ollen neljällä jaolliset erotukset poimitaan alkuperäisestä datasta pitäen aika lineaarisena.

7.3 Synkronian mittaaminen

Mitä tahansa synkroniaa voidaan tutkia vertaamalla havaittujen synnytyspäivien erotusten jakaumaa vastaavaan teoreettiseen nollahypoteesijakaumaan, jossa oletetaan havaintojen olevan riippumattomia eli ei synkronoituneita. Tässä tutkimuksessa aineistoa on niin vähän, että ei kannata tehdä χ^2 -testiä jakaumien vertailemiseksi, vaan verrataan jakaumia pelkästään niiltä osin, jotka ovat merkinä synkroniasta. Yksinkertaisessa synkroniassa siis verrataan, olisiko aineistossa merkitsevästi enemmän nollan mittaisia erotuksia verrattuna riippumattoman tilanteen jakaumaan. Kiimasykli synkroniaa tutkittaessa verrataan lisäksi neljällä jaollisten erotusten osuuksia. Synkronian esiintymistä tutkitaan erikseen kaikille kolmelle otoskoolle eikä tuloksia yhdistetä tilastollisin menetelmin.

Kaikilla otoskoilla menetelmä on kuitenkin pohjimmiltaan sama, sillä vertailtavana on kerrallaan yksi nollahypoteesijakauma ja vastaava empiirinen jakauma. Kahden havainnon otosten aineistolle jakaumia on alunperinkin vain yksi mutta kolmen havainnon otosten aineistolle on erikseen kahden erotuksen ja neljän havainnon aineistolle kolmen erotuksen jakaumat. Jakaumien vertailumenetelmä on yksinkertainen: ajatellaan aineisto binomijakautuneeksi, jossa kukin erotus on synkroniaa edustava to-

dennäköisyydellä P ja ei synkroniaa edustava todennäköisyydellä $1 - P$. Synnytysten synkroniaa tutkittaessa $P = P(0)$ syklisten ajan kaavojen mukaan ja kiimasykliin synkroniaa tutkittaessa kunkin erotuksen todennäköisyys olla neljällä jaollinen lasketaan myöhemmin vastaavasta nollahypoteesijakaumasta. Aineistossa olevalle synkronisten havaintojen osuudelle lasketaan binomijakauman mukainen p -arvo. Testi pidetään yksisuuntaisena, eli jos synkronisten havaintojen osuus on odotettua pienempi, pidetään sitä nollahypoteesin mukaisena tuloksena. Ei siis tulkiteta synkronian vähyyttä millään tavalla.

7.4 Satunnaiset erotukset kolmen ja neljän havainnon välillä

Synnytysten synkroniaa tutkittaessa sovelletaan pienimpien erotusten menetelmää, jolloin varmasti saadaan sisällytettyä analyysiin kaikki nollamittaiset erotukset. Kiimasykliin tutkimista varten sen sijaan ei valita tutkittaviksi pienimpiä erotuksia, sillä ne eivät ole riippumattomia ja täten eivät voi tuottaa yhtä yhteistä p -arvoa tietyn kokoisille otoksille. Nyt kiinnitetään satunnaisesti valittu synnytyspäivä vertailukohdaksi ja verrataan muita havaintoja tähän. Myös tässä menetelmässä jää tutkimatta kolmen havainnon välillä yksi ja neljän havainnon välillä kolme erotusta. Poisjäävät erotukset ovat satunnaisia eivätkä näin ollen aiheuta systemaattista tilastollista harhaa. Toisaalta harrmillista tässä menetelmässä on se, että tulos voi muuttua eri laskentakerroilla riippuen poisjäävästä erotuksesta. Yhdelle kolmen havainnon otokselle olisi siis mahdollista saada kolme erilaista tulosta ja neljän havainnon otokselle neljä, mutta näistä valikoituu satunnaisesti aina yksi. Tästä aiheutuva häiriö pienenee otosmäärän kasvaessa, koska isoilla otosmäärillä poisjäävien erotusten joukossa synkronisten ja ei-synkronisten erotusten suhde lähestyy odotettua arvoaan. Tässä tutkimuksessa aineistoa on käytettävissä tässä mielessä hyvin vähän – kolmen havainnon otoksia on kymmenen ja neljän havainnon otoksia kuusi.

Kolmen ja neljän havainnon välistä kiimasykliin synkroniaa tutkittaessa on neljällä jaollisten erotusten osuutta verrattava satunnaisen tilanteen vastaavaan osuuteen. Lasketaan tämä käyttäen nollahypoteesijakaumaa kahden tasajakautuneen muuttujan erotuksen itseisarvolle. Kun aineistossa kukin satunnaisesti valittu havainto noudattaa tasajakautunutta välillä $[0, T-1]$, kahden havainnon erotuksen itseisarvolle saadaan piste-todennäköisyysjakauma

$$p(|x_a - x_b|) = \begin{cases} \frac{1}{T} & \text{kun } x_a - x_b = 0 \\ \frac{2}{T^2}(T - |x_a - x_b|) & \text{kun } 0 < |x_a - x_b| \leq T - 1 \\ 0 & \text{kun } |x_a - x_b| > T. \end{cases} \quad (7.1)$$

Tästä jakaumasta neljällä jaollisten erotusten osuus on

$$\begin{aligned}
 p(|x_a - x_b| \equiv 0(4)) &= p(|x_a - x_b| = 0) + \sum_{y=1}^{\lfloor T/4 \rfloor} p(|x_a - x_b| = 4y) \\
 &= \frac{1}{T} + \sum_{y=1}^{\lfloor T/4 \rfloor} \frac{2}{T^2} (T - 4y) = \frac{1 + 2\lfloor T/4 \rfloor}{T} - \frac{4\lfloor T/4 \rfloor(\lfloor T/4 \rfloor + 1)}{T^2}. \quad (7.2)
 \end{aligned}$$

Intuitiivisesti neljällä jaollisen erotuksen todennäköisyys rajoittamattomassa lineaarisessa ajassa olisi yksinkertaisesti $\frac{1}{4}$, mutta otetaan tässä huomioon tarkan pistetodennäköisyysjakauman tuoma informaatio rajallisen ajanjakson vaikutuksesta yksittäisiin todennäköisyyksiin. T :n kasvaessa rajatta todennäköisyys lähestyy 0.25:tä, kuten odotettua.

7.5 Kahden havainnon välinen synkronia

Kahden havainnon välillä olevia erotuksia on $44 = n_2$ kpl. Erotukset sekä sykliseen aikaan sovellettuna että sellaisinaan on esitetty histogrammissa 7.1. Esimerkiksi nollearotuksia on molemmissa histogrammeissa kaksinkertainen määrä verrattuna nollahypoteesijakauman mukaiseen odotusarvoon. Lisäksi alkuperäisessä eli ei sykliseen aikaan sovelletussa aineistossa on neljän päivän mittaisia erotuksia juuri odotettu määrä ja 8 ja 12 päivän erotuksia odotettua enemmän. Datassa siis näyttäisi olevan synnytysten synkroniaa ja ehkä myös kiimasykliä synkroniaa.

Kaavasta 3.1 saadaan todennäköisyys nollan pituiselle erotukselle kahden havainnon välillä syklisessä ajassa. Synnytysten synkronian todennäköisyys on sen mukaan $P_{2 \text{ hav}}^{\text{synnytykset}} = P(0)_{2 \text{ hav}} = \frac{1}{T} = \frac{1}{21}$. Vastaava havaittu osuus on $\frac{\text{fr}(0)}{n_2} = \frac{4}{44}$ (ks. kuva 7.1, data syklisessä ajassa). Binomijakaumasta R-ohjelman komennolla `pbinom(3, 44, 1/21, lower.tail=FALSE)` laskettu p-arvo synnytysten synkronialle on 0.156.

Tutkitaan sitten kiimasykliä synkroniaa lineaarisessa ajassa vertaamalla, onko neljällä jaollisia erotuksia merkittävästi enemmän kuin riippumattoman tilanteen jakaumassa. Neljällä jaollisia erotuksia on havaittu $\frac{\text{fr}(0)}{n_2} + \frac{\text{fr}(4)}{n_2} + \frac{\text{fr}(8)}{n_2} + \frac{\text{fr}(12)}{n_2} + \frac{\text{fr}(16)}{n_2} + \frac{\text{fr}(20)}{n_2} = \frac{4}{44} + \frac{2}{44} + \frac{6}{44} + \frac{4}{44} + \frac{1}{44} + \frac{0}{44} = \frac{17}{44}$ (ks. kuva 7.1, alkuperäinen data). Neljällä jaollisen erotuksen todennäköisyys kaavan 7.2 mukaan, kun $T = 21$, on $\frac{11}{21} - \frac{120}{21^2} = 0.252$. Niinpä voidaan laskea p-arvo 17:lle neljällä jaolliselle erotukselle binomijakaumasta $\text{Bin}(44, 0.252)$. Tämä p-arvo on 0.0341.

Kiimasykliä synkronian p-arvo on siis alle 0.05 mutta synnytysten synkronian p-arvo yli 0.1. Kahden havainnon otoksissa on täten 5 %:n riskitasolla kiimasykliä synkroniaa mutta ei havaittavaa synnytysten synkroniaa.

7.6 Kolmen havainnon välinen synkronia

Kolmen havainnon otoksia on $10 = n_3$ kpl. Kaavoista 4.2 ja 4.3 saadaan synnytysten synkronian laskemista varten $d_1:n$ ja $d_2:n$ jakaumat, joista tarvitaan vain nollaerotusten todennäköisyydet. $P(0)_{d_1} = \frac{3T-2}{T^2}$ ja $P(0)_{d_2} = \frac{1}{T^2}$. Havaintoaineistossa on yhteensä kolme nollamittaista otoksen pienintä erotusta eikä yhtään nollan pituista toiseksi pienintä erotusta. P-arvo synnytysten synkronialle $d_1:n$ avulla tarkasteltuna on 0.150 (`pbinom(2, 10, (3*21-2)/21^2, lower.tail=FALSE)`), mikä ei täytä tilastollisen merkitsevyyden ehtoa. $d_2:n$ p-arvoon perustuva tarkastelu ei ole järkevä, sillä todennäköisyys, että datassa on vähintään nolla tiettyä erotusta, on aina 1. Joka tapauksessa, kun synkronisia havaintoja ei ole lainkaan, synkronian olemassaoloa ei ole tarpeen testata tilastollisesti. Synnytysten synkroniaa ei siis voida todeta kolmen havainnon otoksissa pienimmän eikä toiseksi pienimmän erotuksen suhteen.

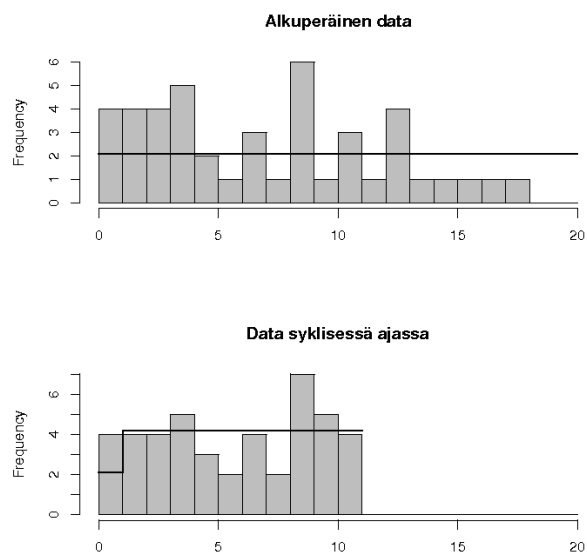
Lasketaan sitten kiimasykliä synkronian tilanne lineaarisessa ajassa. Valitaan kustakin otoksesta satunnainen havainto, johon verrattuna saadaan aina kaksi riippumattomaa erotusta. Yhteiseksi otoskooksi saadaan koko kolmen havainnon otosten aineistolle siten $2 \cdot 10 = 20$. Valitaan nyt otoksista satunnaiset havainnot R-ohjelman komenolla `sample(1:3, 10, replace=TRUE)`, joihin kahta muuta saman otoksen havaintoa sitten verrataan. Saadaan satunnaislukusarja 2 3 1 2 1 3 3 2 3 3. Kun otetaan tämän mukaan kaksi erotusta kustakin otoksesta, näistä 20:stä erotuksesta 10 on neljällä jaollisia. Kuvasta 7.3 erotuksia katsoessa huomaa selvästi korkeat pylväät juuri neljällä jaollisten erotusten kohdalla. Kaavasta 7.2 laskettuna neljällä jaollisen erotuksen todennäköisyys on 0.252. Binomijakaumaan $\text{Bin}(20, 0.252)$ perustuva p-arvo 10 synkroniselle havainnolle on 0.0147, mikä tarkoittaa tilastollisesti merkitsevää synkroniaa kiimasykliä välillä.

7.7 Neljän havainnon välinen synkronia

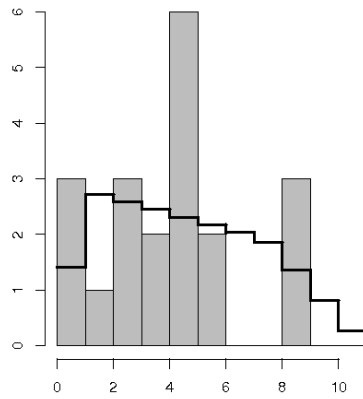
Lasketaan neljä havaintoa sisältävien otosten todennäköisyys osoittaa synnytysten synkroniaa pienimpien, toiseksi pienimpien tai kolmanneksi pienimpien erotuksen perusteella. Käytetään simuloiden saatuja taulukossa 6.1 ilmoitettuja todennäköisyyksiä. Näistä tarvitaan vain nollaerotuksen todennäköisyydet, jotka ovat seuraavat: $P(d_1 = 0) = 0.263$, $P(d_2 = 0) = 0.0156$ ja $P(d_3 = 0) = 0.00888$. Datassa pienimmistä erotuksista 4 on nollia, toiseksi pienimmistä 1 ja kolmanneksi pienimmistä myös 1. Kaikki pienimmät erotukset yhdessä on esitetty histogrammissa 7.4. Esimerkiksi binomijakaumalla $\text{Bin}(6, 0.263)$ saadaan siis p-arvo nollahypoteesille "pienimpien erotusten joukossa ei ole nollaerotuksia enempää kuin riippumattomassa tilanteessa". Pienimmille erotuksil-

le tämä p-arvo on 0.045, toiseksi pienimmille 0.090 ja kolmanneksi pienimmille 0.052. Kaikki p-arvot ovat alle 0.1 eli kaikki kolme osa-alueetta antavat vähintään heikosti merkitsevän tuloksen. Yhteistuloksena voidaan siis sanoa, että neljän havainnon otoksissa on todettavissa heikosti merkitsevää synnytysten synkroniaa.

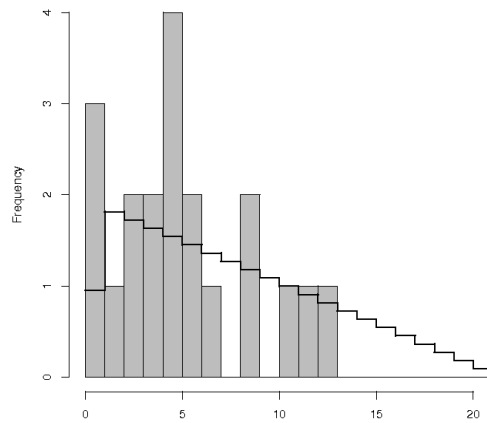
Tutkitaan sitten kiimasykliä synkroniaa lineaarisessa ajassa valiten kustakin otoksesta satunnainen havainto, johon kolmea muuta havaintoa verrataan yksitellen. Neljän havainnon otoksia on vain $6 = n_4$ kpl, mutta tällä menetelmällä saadaan käyttöön kolme riippumatonta erotusta kustakin otoksesta. Kuudessa otoksessa on tällöin yhteensä $6 \cdot 3 = 18$ käytettävissä olevaa erotusta. Satunnaisesti valitut vertailuhavainnot kuudessa otoksessa ovat otoksen sisäisessä havainnointijärjestyksessä 2, 3, 1, 3, 4, ja 4. Näiden mukaan otetuista 18 erotuksesta vain 4 on neljällä jaollisia. Näistä erotuksista on yhteinen histogrammi kuvassa 7.5. Kun neljällä jaollisen erotuksen todennäköisyys on kaavan 7.2 mukaan 0.252, binomijakauman mukainen p-arvo neljälle synkroniselle erotukselle 18:sta on 0.701, mikä ei osoita kiimasykliä synkroniaa.



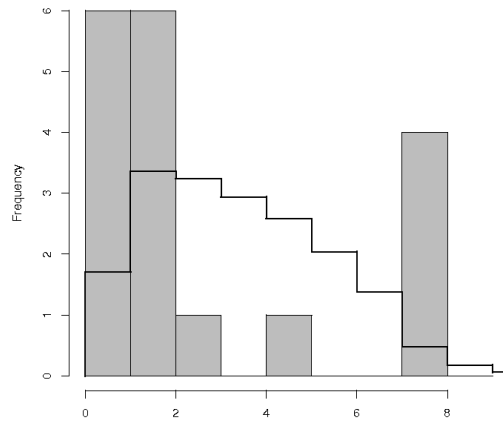
Kuva 7.1: Synnytyspäivien välisten erotusten jakauma kahden havainnon otoksissa datassa (histogrammi) ja nollahypoteesijakaumassa (paksu viiva) sekä alkuperäisessä muodossa että sykliseen aikaan sovelletuna. Kunkin erotuksen frekvenssipylväs on numeron oikealla puolella. (Esim. ensimmäinen pylväs kertoo nollaerotusten määrän.)



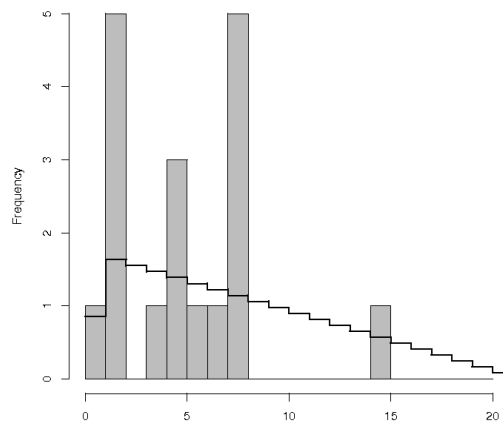
Kuva 7.2: d_1 ja d_2 kolmen havainnon otoksissa ja niiden nollahypoteesijakauma



Kuva 7.3: Kaksi satunnaista erotusta kustakin kolmen havainnon otoksesta ja niiden nollahypoteesijakauma.



Kuva 7.4: Kolme pienintä erotusta d_1 , d_2 ja d_3 neljän havainnon otoksissa ja niiden yhteinen nollahypoteesijakauma.



Kuva 7.5: Satunnaisesti valitut kolme erotusta neljän havainnon otoksissa ja niiden nollahypoteesijakauma.

8 Synkronian tutkiminen aineistoa luokittelemalla

Kolmen ja neljän havainnon otoksista koostuvia aineistoja voidaan tutkia edellisissä luvuissa käytettyä erotusmenetelmää tarkemmin siten, että ei tutkita pelkästään yhdistettyjä $d_{1:n}$, $d_{2:n}$ ja mahdollisesti $d_{3:n}$ jakaumia vaan tutkitaan, minkä havaintojen välillä synkroniaa otoksissa tarkemmin ottaen esiintyy. Kolmen havainnon otoksessa voi olla synkroniaa parin kesken tai kaikkien kolmen havainnon kesken. Neljän havainnon otoksessa taas vaihtoehtoja on neljä: yksi tai kaksi paria on synkroniassa, yksi kolmikko tai koko nelikko on synkroniassa. Näitä tilanteita tarkastelemalla saadaan tarkemmin selville esimerkiksi, onko koko aineistossa oleva synkronia tulosta vain muutamasta vahvasti synkronisesta otoksesta vai onko monissa otoksissa synkroniaa muutamien yksilön välillä. Käytetään menetelmää vain synnytysten synkronian tutkimiseen kolmen havainnon otoksille, sillä kiimasykliä synkronian todennäköisyyksien laskeminen vaatisi paljon lisätyötä. Neljän havainnon otoksille taas nollahypoteesin mukaisten todennäköisyyksien laskeminen on hankalaa; tähän mennessä erotusten jakaumat on määritetty vain simulointien tuloksista.

8.1 Teoria kolmen havainnon otosten luokittelulle

Nollahypoteesin mukaiset todennäköisyydet synkronialuokille lasketaan suurimmaksi osaksi teoriassa aiemmin saatujen tulosten pohjalta. Kolmen havainnon otoksille on laskettava kolme todennäköisyyttä, jotka summautuvat ykköseksi:

$P(\text{mitkään havainnot eivät synkroniassa}) = P(0/3)$, $P(\text{kaksi havaintoa synkroniassa}) = P(2/3)$ ja $P(\text{kaikki kolme havaintoa synkroniassa}) = P(3/3)$.

Käytetään syklistä ajankäsitystä siten, että tutkitaan havaintojen välillä olevia pienimpiä erotuksia sekä niiden summaa peilaamattomana. Tämä siksi, että aikaa on hankala käsitellä täysin jatkuvana syklinä, jos samalla haluttaisiin tietää tarkkaan, mitkä havainnot ovat kiimasykliä suhteen synkroniassa. Jos esimerkiksi havainnot ovat $x_1 = 0$, $x_2 = 4$ ja $x_3 = 12$ ja ajatellaan aika alkavaksi nolasta, kaikki kolme peräkkäistä erotusta ovat neljällä jaollisia. Jos taas ajatellaan aika alkavaksi toisen havainnon kohdalta, vain kaksi havaintoa ovat synkroniassa, $x_2 = 4$ ja $x_3 = 12$, sillä väli x_3 :sta x_1 :een ei ole neljällä jaollinen. Jos T on jaollinen neljällä niin ongelmaa ei ole, kun kiimasyklin merkinä pidetään neljällä jaollista synnytyspäivien erotusta. Aineistossa T on kuitenkin 21, joten valitaan tutkittavaksi kaksi pienintä erotusta ja niiden peilaamaton

summa, joka on maksimissaan $\frac{5}{6}T$. Tällöin oletetaan vahvasti, että myyrät synnyttävät jatkuvasti 21 päivän välein ja synkronia ilmenee parhaiten lähekkäin tapahtuneiden synnytysten välillä. Ilman sykliysysoletusta valittaisiin aina alkuhetkeksi ensimmäisenä tapahtunut synnytys ja verrattaisiin toista ja kolmatta synnytystä siihen.

8.2 Synkronialuokkien todennäköisyyksien laskeminen

Nyt voidaan määrittää luokittaiset todennäköisyydet kullekin synkronialuokalle kolmen havainnon otoksia varten. Lyhennetään ilmaisu ” d_i on synnytysten suhteen synkroninen erotus” merkinnäksi $s_s(d_i)$ ja vastaavasti kiimasykliin synkronian suhteen merkinnäksi $s_{ks}(d_i)$. Se, että erotus ei ole synkroniaa edustava, merkitään $\bar{s}_s(d_i)$ tai $\bar{s}_{ks}(d_i)$. Synnytysten synkronian merkinä on nollierotus ja kiimasykliin synkronian merkinä neljällä jaollinen erotus, jota merkitään lyhennetyllä modulo-ilmaisulla $d_i \equiv 0(4)$. Laskuja puretaan auki käyttäen todennäköisyyslaskennan sääntöä $P(A \cap B) = P(A) \cdot P(B | A)$. Kiimasykliin liittyvistä laskuista esitetään vain alkuosa, loppuun asti laskeminen vaatisi paljon lisätyötä. Lasketaan erikseen synnytysten ja kiimasykliin suhteen todennäköisyydet sille, että kaikki kolme havaintoa ovat synkroniassa.

$$\begin{aligned} P_s(3/3) &= P(s_s(d_1), s_s(d_2), s_s(d_1 + d_2)) \\ &= P(s_s(d_1), s_s(d_2)) \cdot P(s_s(d_1 + d_2) | s_s(d_1), s_s(d_2)) \\ &= P(d_1 = 0, d_2 = 0) \underbrace{P(d_1 + d_2 = 0 | d_1 = 0, d_2 = 0)}_{=1} \\ &= P(d_2 = 0) \cdot \overbrace{P(d_1 = 0 | d_2 = 0)}^{=1, \text{ sillä } d_1 \leq d_2} = \frac{1}{T^2} \end{aligned}$$

$$\begin{aligned} P_{ks}(3/3) &= P(s_{ks}(d_1), s_{ks}(d_2) \text{ ja } s_{ks}(d_1 + d_2)) \\ &= P(d_1 \equiv 0(4), d_2 \equiv 0(4)) \cdot \underbrace{P(d_1 + d_2 \equiv 0(4) | d_1 \equiv 0(4), d_2 \equiv 0(4))}_{=1, \text{ neljällä jaollisten lukujen summa on jaollinen neljällä}} \\ &= P(d_1 \equiv 0(4), d_2 \equiv 0(4)) \end{aligned}$$

Lasketaan seuraavaksi todennäköisyys sille, että kaksi havaintoa kolmesta on synkroniassa.

$$\begin{aligned}
P_s(2/3) &= P(s_s(d_1) \text{ ja } s_s(d_2)) + P(s_s(d_1) \text{ ja } s_s(d_2)) + P(s_s(d_1), s_s(d_2) \text{ ja } s_s(d_1 + d_2)) \\
&= P(d_1 = 0, d_2 \neq 0) + P(d_1 \neq 0) \underbrace{P(d_2 = 0 \mid d_1 \neq 0)}_{=0, d_1 \leq d_2} \\
&\quad + P(d_1 \neq 0, d_2 \neq 0) \underbrace{P(d_1 + d_2 = 0 \mid d_1 \neq 0, d_2 \neq 0)}_{=0} \\
&= P(d_1 = 0) - P(d_1 = 0, d_2 = 0) \quad \parallel P(A \cap B^C) = P(A) - P(A \cap B) \\
&= \frac{3T - 2}{T^2} - \frac{1}{T^2} = \frac{3T - 3}{T^2} \\
P_{ks}(2/3) &= P(s_{ks}(d_1), s_{ks}(d_2) \text{ ja } s_{ks}(d_1 + d_2)) + P(s_{ks}(d_1), s_{ks}(d_2) \text{ ja } s_{ks}(d_1 + d_2)) \\
&\quad + P(s_{ks}(d_1), s_{ks}(d_2) \text{ ja } s_{ks}(d_1 + d_2)) \\
&= P(d_1 \equiv 0(4), d_2 \not\equiv 0(4)) + P(d_1 \not\equiv 0(4), d_2 \equiv 0(4)) \\
&\quad + P(d_1 + d_2 \equiv 0(4) \mid d_1 \not\equiv 0(4), d_2 \not\equiv 0(4)) \cdot P(d_1 \not\equiv 0(4), d_2 \not\equiv 0(4))
\end{aligned}$$

Viimeiseksi lasketaan todennäköisyys sille, että kolmikossa ei esiinny ollenkaan synkroniaa. Lasketaan osa laskuista komplementin todennäköisyyden kautta, jolloin voidaan käyttää apuna aikaisempia laskuja.

$$\begin{aligned}
P_s(0/3) &= P(s_s(d_1), s_s(d_2) \text{ ja } s_s(d_1 + d_2)) \\
&= P(s_s(d_1), s_s(d_2)) \cdot P(s_s(d_1 + d_2) \mid s_s(d_1), s_s(d_2)) \\
&= P(d_1 \neq 0, d_2 \neq 0) \cdot \underbrace{P(d_1 + d_2 \neq 0 \mid d_1 \neq 0, d_2 \neq 0)}_{=1, \text{ positiivisten lukujen summa on yli nollan}} \\
&= P(d_1 \neq 0) \underbrace{P(d_2 \neq 0 \mid d_1 \neq 0)}_{=1, d_1 \leq d_2} = P(d_1 \neq 0) = 1 - \frac{3T - 2}{T^2} \\
P_{ks}(0/3) &= P(s_{ks}(d_1), s_{ks}(d_2) \text{ ja } s_{ks}(d_1 + d_2)) \\
&= P(s_{ks}(d_1), s_{ks}(d_2)) \cdot P(s_{ks}(d_1 + d_2) \mid s_{ks}(d_1), s_{ks}(d_2)) \\
&= P(s_{ks}(d_1), s_{ks}(d_2)) \cdot \left(1 - P(s_{ks}(d_1 + d_2) \mid s_{ks}(d_1), s_{ks}(d_2))\right) \\
&= P(d_1 \not\equiv 0(4), d_2 \not\equiv 0(4)) \underbrace{\left(1 - P(d_1 + d_2 \equiv 0(4) \mid d_1 \not\equiv 0(4), d_2 \not\equiv 0(4))\right)}_{\text{saadaan laskusta } P_{ks}(2/3)}
\end{aligned}$$

Synnytysten synkronialuokkien yhteinen todennäköisyys on yksi, kuten pitääkin:

$$\frac{1}{T^2} + \frac{3T - 3}{T^2} + 1 - \frac{3T - 2}{T^2} = 1.$$

8.3 Luokittelumenetelmän soveltaminen aineistoon

Numeeriset arvot teoreettisille synnytysten synkronian luokittaisille todennäköisyyksille sekä havaintoaineistossa olevat vastaavat osuudet kymmenen otoksen kesken, kun T on 21, ovat

$$\begin{aligned} P_s(3/3) &= \frac{1}{T^2} = \frac{1}{441} \approx 0.00227 & \frac{\text{fr}(3/3)}{10} &= 0 \\ P_s(2/3) &= \frac{3T-3}{T^2} = \frac{60}{441} \approx 0.136 & \frac{\text{fr}(2/3)}{10} &= \frac{3}{10} = 0.3 \\ P_s(0/3) &= 1 - \frac{3T-2}{T^2} = 1 - \frac{61}{441} \approx 0.862 & \frac{\text{fr}(0/3)}{10} &= \frac{7}{10} = 0.7. \end{aligned}$$

Kolmessa otoksessa kymmenestä on synkroninen pari eikä yhdessäkään ole synkronista kolmikkoa. Tämä erittely ilmentää sen, että synkronia ilmenee kolmen havainnon otoksessa aina vain kahden myyrän välillä, ei kertaakaan koko kolmikossa. Tilanteen p -arvo kolmiluokkaisen multinomijakauman perusteella on 0.107 (`dmultinom(c(0,3,7),10,c(1,60,380)/441)`). Kolmen havainnon otoksissa ei siis ilmene nollahypoteesista tilastollisesti merkitsevässä määrin poikkeavaa synkronian jakautumista otosten sisällä.

9 Johtopäätökset

Tässä tilastotieteen syventäviin opintoihin liittyvässä opinnäytetyössä on kehitetty muutamia erilaisia menetelmiä synkronian mittaamiselle syklisessä tai lineaarisessa ajassa. Menetelmiä on sovellettu dataan, joka on saatu rekisteröimällä luonnossa lähekkäin elävien ja yhtä aikaa tiineenä olleiden metsämyyränaaraiden synnytyspäivät. Yhteensä kahden, kolmen ja neljän myyrän otoksia on 60. Tässä luvussa esitetään saadut tulokset kahden eri synkronialaadun, synnytysten ja kiimasykliä synkronian suhteen. Synnytysten synkronia tarkoittaa sitä, että tietyn otoksen myyrät ovat synnyttäneet samana päivänä (tai 21 päivän välein, jota tosin ei datassa ilmennyt) ja kiimasykliä synkronia taas sitä, että myyrien neljän päivän välein esiintyvä kiima esiintyy aina samana päivänä. Tämä ilmenee siten, että synnytysten välillä on neljällä jaollinen määrä päiviä.

Taulukossa 9.1 on esitetty tulokset kahden, kolmen ja neljän havainnon otoksille sekä synnytysten että kiimasykliä synkronian suhteen ilmoittamalla vastaavaan nollahypoteesiin liittyvä p-arvo. Nollahypoteesin mukainen tilanne on se, että otoksissa ei esiinny lainkaan kyseistä synkroniaa vaan synnytykset jakautuvat satunnaisesti ja muista myyristä riippumattomasti. Synnytysten synkroniaa laskettaessa aika on käsitelty 21 päivän pituisena syklinä, kiimasykliä synkroniaa laskettaessa taas lineaarisena. Synnytysten synkroniaa tutkittaessa on tutkittu pienimpiä erotuksia: 3 havainnon otoksissa kahta ja 4 havainnon otoksissa kolmea pienintä erotusta. Tällöin on saatu 2 havainnon otoksille yksi p-arvo, 3 havainnon otoksille kaksi ja 4 havainnon otoksille kolme, sillä nämä erotukset eivät ole toisistaan riippumattomia. Kiimasykliä synkroniaa tutkittaessa taas on verrattu satunnaisesti valittua havaintoa yksitellen muihin havaintoihin, jolloin erotukset ovat riippumattomia ja niille saadaan yksi yhteinen p-arvo.

Otoskoko	Synnytysten synkronia	Kiimasykliä synkronia
2 havaintoa	0.16	0.03 **
3 havaintoa	0.15 ja 1	0.01 **
4 havaintoa	0.045, 0.09 ja 0.05 *	0.70

Taulukko 9.1: P-arvot synkronialle otoskoon ja synkronian laadun mukaan

P-arvoihin liittyvät *-merkinnät kuvaavat tuloksen tilastollista merkitsevyyttä: yhden tähden tulos on tilastollisesti merkitsevä 90 %:n luottamustasolla, kahden tähden

tulos 95 %:n ja kolmen tähden tulos 99 %:n luottamustasolla.

Taulukosta 9.1 nähdään, että kaiken kokoisissa otoksissa on havaittavissa jompaakumpaa synkroniaa ja että synkroniaa esiintyy sekä synnytysten että kiimasykliin suhteen. Synnytysten synkroniaa ilmenee vain neljän havainnon otoksissa, ja näissä otoksissa populaatio on oletettavasti tiheämpi kuin muissa. Synkronian ilmeneminen erityisesti tiheissä populaatioissa on ilmiö, jota on pohdittu aiemminkin ja jolle on olemassa biologisia perusteita. (Ks. luku 2) Kiimasykliin synkronia taas ei noudata vastaavaa sääntöä vaan näyttää ilmenevän erityisesti harvalukuisissa populaatioissa.

Erityisesti neljän havainnon otosten aineisto poikkeaa mielenkiintoisen näköisesti nollahypoteesin tilanteesta. (ks. kuva 7.5). Kun sadan neliömetrin alueelta on saatu neljä naarasta, populaatio on oletettavasti tiheä. Johdannossa käsiteltyjen artikkelien perusteella mahdollinen synkronoitumisilmiö (oletettavasti epäsynkronia) olisi selvimmän näkyvässä nimenomaan tiheissä populaatioissa. Tässä tutkimuksessa ei kuitenkaan käsitellä epäsynkroniaa muuten kuin kiimasykliin synkronian kautta. Mikäli kiimasykliin pituutena pidetään neljää päivää, kiimasykliin synkroniaa ei neljän havainnon otoksissa ole lainkaan, mutta kolmen ja puolen päivän sykli voisi olla biologisesti mahdollinen, ja sille neljän havainnon otosten data antaisi erittäin vahvan tuen. Koska synnytysten synkronia ilmenee joka tapauksessa vahvemmin tiheämissä populaatioissa, voisi muissa tutkimuksissa olla hyödyllistä tutkia 3,5–4 päivän pituisen kiimasykliin synkronian/epäsynkronian esiintymistä tiheissä myyräpopulaatioissa.

Käytettävissä olevan havaintoaineiston perusteella nähdään, että luonnonoloissa elävien metsämyyräiden kesken esiintyy jonkin verran sekä synnytysten että neljän päivän pituisen kiimasykliin synkroniaa. Toisaalta puolesta käsitellyistä tapauksista synkronian olemassaoloa ei voitu käytetyin tilastollisin menetelmin todeta. Ainakin synnytysten synkronia näyttäisi olevan vahvempi ilmiö tiheämissä populaatioissa. Samoin myös kiimasykliin synkronia on vahvimmillaan tiheissä populaatioissa, jos kiimasykliin pituus on 3,5 päivää, mitä tässä tutkimuksessa tosin ei ole analysoitu tarkemmin.

A Aineisto

Aineisto koostuu kahden, kolmen ja neljän havainnon kokoisista otoksista. 2 havainnon otoksia on 44, 3 havainnon otoksia 10 ja 4 havainnon otoksia 6. Ilmoitetaan kustakin otoksesta synnytyspäivät ensimmäisestä raportointipäivästä lähtien.

	2 hav.	3 havaintoa	4 havaintoa
1	11 14	3 5 8	4 11 11 18
2	15 15	7 15 18	5 5 6 9
3	3 16	11 15 17	2 9 9 16
4	14 14	7 11 16	8 14 15 16
5	15 18	2 10 14	10 10 10 14
6	2 19	12 17 17	9 10 14 15
7	5 16	4 8 12	
8	1 16	7 7 11	
9	3 13	14 15 15	
10	4 13	7 15 17	
11	1 3		
12	8 10		
13	9 10		
14	6 14		
15	8 8		
16	12 18		
17	7 15		
18	3 11		
19	6 14		
20	15 15		
21	9 13		
22	9 12		
23	14 16		
24	9 10		

	2 hav.	3 havaintoa	4 havaintoa
25	9	15	
26	5	17	
27	9	15	
28	12	15	
29	5	15	
30	15	17	
31	5	17	
32	10	13	
33	4	11	
34	3	19	
35	6	14	
36	3	15	
37	10	15	
38	5	17	
39	7	11	
40	7	17	
41	6	7	
42	15	16	
43	2	16	
44	8	16	

B Kolmen havainnon otosten simulointiin käytetty R-koodi

```
sim.3hav<-function(T,n,N) {
# Simuloidaan N simuloitua havaintoaineistoa
# Kukin havaintoaineisto sisältää n kpl 3 havainnon otoksia
# Otoksista saadaan tietää kaksi pienintä erotusta
# T: syklin maksimi
# N: simulaatioiden määrä
# n: otosten määrä
aineistot<-matrix(nrow=N,ncol=2*n)
for (i in 1:N) aineistot[i,]<-s12s(T,n)
# aineistot-matriisin rivi sisältää n kpl 3 havainnon otoksia
# otos: c(d1,d2)
aineistot
}

s12s<-function(T,n) {
# simuloidaan 3 havainnon otoksille kahden pienimmän erotuksen jakaumat
# T syklin maksimi
# n otosten määrä
arvot<-sample(0:(T-1),3*n,replace=TRUE)
#jaetaan arvot kolmesarakkeiseen matriisiin
arvot<-matrix(arvot,ncol=3)
# otetaan talteen 2 pienintä erotusta 3 havainnon välillä
erotukset<-deeksi3pienet(T,arvot[,1],arvot[,2],arvot[,3])
erotuksetPeilattu<-peilaa(T,erotukset)
erotusVektorit<-NULL
for(i in 1:n)
erotusVektorit<-c(erotusVektorit,erotukset[i,])
erotusVektorit
}

deeksi3pienet<-function(T,x1,x2,x3) {
```

```

# tuottaa 2 pienintä erotusta 3 havainnon välillä
syklisen ajankäsityksen mukaan
# kukin xi-vektori sisältää otosten i. havainnot
# otosten lukumäärä ilmenee x-vektorien pituutena (kaikki yhtä pitkiä)
d<-abs(c(x3-x2,x2-x1,x3-x1))
# peilataan havainnot noudattamaan syklistä ajankäsitystä
p1<-T-d*(d>floor((T-1)/2))
p2<-d[d<=floor((T-1)/2)]
p<-replace(p1,p1==T,p2)
# kootaan kunkin otoksen tiedot omalle rivilleen
d.kaikki<-matrix(p,byrow=FALSE,nrow=length(x1))
# järjestetään rivi nousevaan järjestykseen
d.jarj<-apply(d.kaikki,1,sort)
d.jarj<-matrix(d.jarj,byrow=TRUE,nrow=length(x1))
# valitaan riviltä 2 ensimmäistä eli pienimmät alkiot
d<-d.jarj[,1:2]
}

```

Viitteet

- [1] Rolf Anker Ims, Soren Bondrup-Nielsen and Chr. Stenseth 1988, *Temporal patterns of breeding events in small rodent populations*, OIKOS 53:2, sivut 229-234.
- [2] Edda Johannesen, Harry P. Andreassen and Rolf A. Ims 2000, *The effect of patch isolation on reproductive synchrony in the root vole*, OIKOS 89:1, sivut 37-40.
- [3] Xavier Lambin 1993, *Determinants of the synchrony of reproduction in Townsend's voles, Microtus townsendii*, OIKOS 67:1, sivut 107-113.
- [4] William McShea 1989, *Reproductive synchrony and home range size in a territorial microtine*, OIKOS 56:2, sivut 182-186.
- [5] Tanja Poikonen, Esa Koskela, Tapio Mappes and Suzanne C. Mills 2007, *Infanticide in the evolution of reproductive synchrony: Effects on reproductive success*, Evolution (27.11.2007 vielä julkaisematon).
- [6] R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org> (viitattu 27.11.2007).

Kuvat

1.1 Metsämyyrä Matti Laineen kuvaamana	2
3.1 Sykli, jossa $T = 20$	5
4.1 Epäsynkronian maksimi parittomassa syklissä.	22
4.2 Epäsynkronian maksimi parillisessa syklissä.	22
4.3 fr_{d_1} :n määrittämiseen liittyvä apukuva	23
4.4 fr_{d_1} :n määrittämiseen liittyvä apukuva	23
4.5 fr_{d_1} :n määrittämiseen liittyvä apukuva	23
4.6 fr_{d_1} :n määrittämiseen liittyvä apukuva	24
4.7 fr_{d_1} :n määrittämiseen liittyvä apukuva	24
4.8 fr_{d_1} :n määrittämiseen liittyvä apukuva	24
4.9 fr_{d_1} :n määrittämiseen liittyvä apukuva	25
4.10 fr_{d_2} :n määrittämiseen liittyvä apukuva	25
4.11 fr_{d_2} :n määrittämiseen liittyvä apukuva	26
4.12 fr_{d_2} :n määrittämiseen liittyvä apukuva	26
4.13 fr_{d_2} :n määrittämiseen liittyvä apukuva	27
4.14 fr_{d_2} :n määrittämiseen liittyvä apukuva	27
4.15 fr_{d_2} :n määrittämiseen liittyvä apukuva	27
4.16 fr_{d_2} :n määrittämiseen liittyvä apukuva	28
4.17 fr_{d_2} :n määrittämiseen liittyvä apukuva	28
5.1 d_1 :n ja d_2 :n teoreettinen ja simuloitu jakauma	31
5.2 d_1 :n simuloitujen jakauman tarkastelua	32
6.1 Simuloitu jakauma neljän havainnon otoksille	35
7.1 Kahden havainnon otokset	42
7.2 Kolmen havainnon otosten kaksi pienintä erotusta	43
7.3 Satunnaiset erotukset kolmen havainnon otoksista	43
7.4 Neljän havainnon otosten kolme pienintä erotusta	44
7.5 Satunnaiset erotukset neljän havainnon otoksista	44